Description and Necessity: Towards a Cognitive Science of Word Meaning

Nicholas Richard Braisby

Ph.D.
University of Edinburgh
1990

Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself unless otherwise indicated.

Edinburgh, 29th May 1990.

Nick Braisby

Acknowledgements

This thesis would not have been possible were it not for a whole host of people to whom I owe an inestimable debt of gratitude. Firstly, I thank my supervisors.

Robin Cooper has so often, and with such patience, listened to my inane ramblings and, instead of treating me with the contempt I deserved, listened and made far more sense of things than I. Were it not for his constant help, attention, criticism, support, encouragement and, most importantly, friendship, I would probably have given up long ago. If anything fruitful comes of this piece of work, it will mostly be due to him.

Keith Stenning and Terry Myers have provided numerous cogent comments and helpful stimulation. If the research reported herein does not come up to psychological scratch, it is a reflection on my own failings as a student rather than their ability to supervise.

The Centre for Cognitive Science in Edinburgh has been a wonderful place in which to work and my time here has been marked by tremendous pleasure. To all the staff and students go my warmest thanks. But, to Betty, goes a very special thankyou: you're the jewel in the crown. Special thanks too to Jerry Seligman for his patient and considerable help with the formatting of this monstrosity. You helped me get to the binders on time.

I owe a greater debt of gratitude to my very dearest friends. How they kept me on the rails, I can only guess. To Ashoke, Paul, Liz, Richard, Brad and Netty: your warmth and kindness has given me such tremendous happiness over the years that my debt is incalculable. It won't prevent me from trying to repay it. You have never strayed for long from my thoughts, nor will you ever. A special thankyou to Brad, Netty, Richard and Ashoke for supporting me when I needed it most. You taught me how to be myself and to be happy with it, a lesson that can only be learnt from such dear friends as you. And to Jonathan, a special thankyou too. You kept me on the straight and narrow, though not in every sense, of course. You have given me joy that I will never forget.

Finally, I thank my family. Mum, Dad, Martin, Jonathan, Nicola and now, wonderfully,

Lydia. You have provided me with support, strength, hope, courage, energy, encouragement, friendship and love. You have tolerated more than I can imagine, certainly more than I would hope to admit. Your acceptance has given me such a feeling of well-being and purpose that, to describe my feelings would take more than this thesis. My debt to you all is enormeous and my resources so meagre. Of things substantial, I offer you all that I have: my thanks and my love.

Abstract

In this thesis we present a cognitive-scientific view of word meaning. We begin by spelling out some general enterprises of which this thesis may form a small part. Enterprises such as the entwining of facts concerning the epistemology of language use and semantic analysis; the orientation of a study of word meaning as part of a study of thought; and the restoration of a study of thought to the psychological domain. We also indicate some of the tools we employ in the rest of the thesis.

One tool is that provided by Situation Theory and particularly its notion of conditional constraints. Another is a pre-theoretic point that encourages to distinguishes certain semantical questions from related ontological ones. Another tool is a set of puzzles of word meaning an analysis of which forms one objective. These and other tools allow us to isolate some points at which theories of sense and word meaning may diverge. And this prepares the ground for a discussion of competing theories.

Our discussion is informed by psychological concerns and it is these that we take to motivate a general account of word meaning, Sense Generation, and a specific version of this, the Relational View. Sense Generation, we argue respects linguistic constraints concerning ambiguity as well as psychological constraints concerning concepts. We argue, further, that the details of the Relational View are supported by psychological argumentation. And we meet our first objective which is to present an analysis of the various puzzles.

We then turn to the philosophical literature in order to combat some well established positions and arguments which count against the efficacy of Sense Generation views. Sense Generation views hold that the meanings of words are descriptional; opposing this view are the arguments of Kripke and Putnam. We consider first the literature on proper names whose meanings, it is argued, are not descriptional. Our position with respect to this argument is that it be treated as an inadvertent *reductio* of the traditional conception of analyticity. We then turn our attention to the psychological literature on natural kind concepts and suggest that these do not standardly address

the problems that Kripke and Putnam raise. Again, our suggestion is that with an alternative conception of analyticity, one resulting from Situation Theory and Sense Generation, the Kripke-Putnam argument concerning natural kinds can similarly be viewed as a *reductio*. We spell out the Relational View of natural kind concepts.

Our attention then turns to a major psychological theory of concepts, prototype theory. We argue that it fails to meet various constraints of a psychological nature and, further, that these constraints are met by the Relational View, which we re-label the Family of Constraints view. We attempt to adduce reasons for prototype theory's shortcomings by giving a detailed comparison between it and the Family of Constraints view.

Contents

Chapter 1

The Background

> Two essentially different modes of analysis are possible in the study of psychological structures. It seems to us that one of them is responsible for all the failures that have beset former investigators of the old problem, which we are about to tackle in our turn, and that the other is the only correct way to approach it.
>
> The first method analyses complex psychological wholes into *elements*. It may be compared to the chemical analysis of water into hydrogen and oxygen, neither of which possess the properties of the whole and each of which possesses properties not present in the whole. The student applying this method in looking for the explanation of some property of water — why it extinguishes fire, for example — will find to his surprise that hydrogen burns and oxygen sustains fire. These discoveries will not help him in solving the problem. Psychology winds up in the same kind of dead end when it analyses verbal thought into its components, thought and word, and studies them in isolation from each other. In the course of analysis, the original properties of verbal thought have disappeared. Nothing is left to the investigator but to search out the mechanical interaction of two elements in the hope of reconstructing, in a purely speculative way, the vanished properties of the whole.
>
> Psychology, which aims at a study of complex holistic systems, must replace the method of analysis into elements with the method of analysis into units. What is the unit of verbal thought that is further unanalyzable and yet retains the properties of the whole? We believe that such a unit can be found in the internal aspect of the word, *word meaning.*
>
> *(Vygotsky, 1962; p. 5)*

This piece from Vygotsky's *Thought and Language* represents a part of Vygotsky's at-

tempt to spell out his research programme in investigating the nature of verbal thought. It also serves as a good indication of the position which we adopt in this thesis. In moving toward a cognitive science of word meaning, our aims are threefold: to establish some framework for describing word meaning; to account for much of what is known concerning the psychology of concepts; and to suggest the re-orientation of both word meaning and concepts as central aspects of a cognitive scientific study of thought.

Vygotsky's own observations concerning the nature of verbal thought are startling in their breadth. Indeed, we will use some of the examples he cites later in this thesis. However, there is one aspect of Vygotsky's view of word meaning to which we do not subscribe, his apparently internalist stance: a word meaning is an "internal aspect". One would think that this commits Vygotsky to a version of what we may, following Putnam, call the "meanings are in the head" hypothesis. Whether or not this is the case, as will become apparent, this hypothesis is not one we subscribe to. Indeed, according to the framework we develop later, meaning relations are very much external.

The force of Vygotsky's claim, however, is much broader than any debate between internalist and externalist view of meaning. If true, his suggestion is that in studying verbal thought, the twin aspects of words and thoughts are inseparable. To study one in isolation from another is analogous to studying hydrogen separately from oxygen: such an enterprise will never licence valid claims concerning the macroscopic properties of water. Similarly, to study words in isolation from thoughts will never be informative as regards verbal thought.

Surprisingly, perhaps, support for this general claim can be seen in Frege. The later Frege's concerns were such as to centrally locate the fact of language being cognitively significant. Accordingly, rather than viewing expressions of language as possessing just a Meaning or referent, he held that they also possessed a sense, and that therein lay the expression's cognitive significance. Frege viewed the individuation of senses as deriving from propositional-attitude psychology, hence Evans' (1976) suggestion that senses are to be equated with "ways of thinking". Whether this controversial reading of Frege is correct is not a matter for this thesis. That Frege introduces cognitive significance

into semantic theory is. Apparently, then, the enterprises in which Frege and Vygotsky locate themselves should not be seen as too distinct. This, unfortunately, is not true of their respective literatures today.

Current psychological theories of concepts are disparate. In this thesis we will not focus on all, but on the traditional analytic view of concepts and that which purports to replace it, prototype theory. Both views encounter difficulties, but the literature is confused as to the *role* of these theories of concepts. Traditional analytic views, as we shall see, can require the specification of conceptual content to lie beyond the possible apprehension of the possessors of concepts. Prototype views, while apparently accounting for much of what possessors of concepts are deemed to know in virtue of their possession of those concepts, seem unable to account for aspects of thought in the way that was required of the traditional analytic view. Accordingly, some subscribe to hybrid or binary views of concepts: concepts contain cores which are compositional and the domain of thought; they also contain identification procedures which give rise to prototype effects. Such subscribers ordinarily hold that conceptual cores behave according to the traditional analytic view. Here, the problems of the two separate views are confounded, for this last suggestion is such as to presume that the contents of thought may lie beyond the possible apprehension of possessors of those thoughts and their constituent concepts. Psychological experimentation may reveal the nature of identification procedures, but it does not reveal the content of thought.

Current philosophical views on word meaning, and here we are thinking primarily of the views of Kripke and Putnam, are predominated by the view that many words simply do not have Fregean senses. The argument is advanced most forcefully in the case of natural kind terms and proper names but both Kripke and Putnam appear committed to the view that the argument can be applied to many other words. Putnam, in *Is Semantics Possible?*, suggests the following.

> A theory which correctly describes the behaviour of perhaps three hundred words has been asserted to correctly describe the behaviour of tens of thousands of general names.
>
> *(Putnam, 1975; p. 141)*

The point that we shall attempt to make during this thesis is that if it is true that general names do not, in general, possess Fregean senses then the possibility of a theory of semantics ever explaining *uses* of words is limited at best. By rejecting Fregean senses, out of which are composed thoughts, we also reject the motivation behind the notion of sense, which was to explain the cognitive significance of language. For it is in the sense of an expression that its cognitive significance, its very contribution to thought, is deemed to lie. Rejected, also, is the view that propositional-attitude psychology should play a role in a theory of semantics. For it is the attitudes that one may adopt towards a sentence which Frege seemed to suggest lay behind his individuation criterion for senses. Since Frege's notion of sense, then, is so intimately connected to an account of thoughts, the rejection of the Fregean position also amounts to a rejection of Vygotsky's approach.

A consequence of the Kripke-Putnam position is that it distances a study of word meaning from what may be empirically observed concerning the attitudes language users may take towards uses of these words. That is, the meaning of a word is deemed to be independent of the attitudes one may take towards its uses. One goal of this thesis is, then, to suggest that this bifurcation between word and thought is not a necessary one. I hope to suggest that we have been led astray by an insistence on a particular characterisation of concepts and of word meaning, one inherited from a traditional analytic conception of meaning. I hope to suggest that there are real choices that cognitive scientists may take in approaching a study of word meaning and I wish to argue for the efficacy for one set of such choices. Having done so, I will attempt to show that this set of choices does, indeed, allow for the bringing together of empirical and theoretical studies of word and thought.

In this chapter, our objective is to make explicit those aspects of a theory of word meaning which will be treated as implicit throughout the later chapters. We begin by drawing a rarely made distinction between a semantic question, concerned with the conditions under which a word *X* may be used, and an ontological question, concerned with what an X actually *is*. In certain respects, this simple distinction can be viewed as one of the main concerns of the thesis. For it is the making of this distinction that frees us from many traditional restrictions on constructing a framework to describe

word meaning. So, having made the distinction, we turn to some puzzles of word meaning: an analysis will be postponed until Chapter 4. However, in the following section, we establish some intuitions concerning these puzzles and those aspects of the psychological literature where, it seems, answers might be forthcoming. We then turn to consider some of the formal apparatus that we will employ during the rest of the thesis. This apparatus weds our approach of word meaning to a view of information and, in the following section, we pre-empt a naive interpretation of this wedding. Finally, we preview the rest of the thesis and lay down the skeleton of the arguments around which we later attempt to lay some flesh.

1.1 Drawing a Distinction

In some respects, one of the main claims made in this thesis regards the importance of distinguishing questions of semantics from questions of ontology. This is not to say that these questions are independent, but simply that they are separate. With respect to word meaning, throughout the thesis, we will rely on being able to distinguish two claims associated with the use of words: one to do with the information uses can carry and one to do with the ontological commitments such uses make. Our contention will be that a semantic theory should be concerned more with the former than the latter. Let us consider an example.

Suppose *noj* is a fully-fledged word: that is, it plays a full role in normal language use. It is uncontroversial to assume, then, that attached to this word, there are certain conditions of use, call them *noj-conditions*. Now these conditions are such that, provided they obtain, a *use* of *noj* is licenced by the grammar of the language, taking grammar in a wide sense. That is, such conditions are taken to be implicated in determining when a given word may be used. Now it is also uncontroversial to assume that, provided these noj-conditions obtain, it is their obtaining that invest uses of *noj* with informational content. That is, uses of *noj* can convey information to the effect that noj-conditions obtain. Of course, uses may be infelicitous, they may be used in cases where noj-conditions do not obtain, and such uses provide cases of misinformation. But, in general,

a use of *noj* will convey information concerning the obtaining of noj-conditions.

Now the above description suffices as a picture of what this thesis attempts to do. For various words, at various points in the thesis, we will try and explicate a theory of their conditions of use, one which does justice to a number of problematic observations. However, while it is clear that such an objective is uncontroversial from the point of view of studying word meaning, before we begin we must be equally clear about what it is we are *not* trying to do.

A nonsense word such as *noj* illustrates well a discussion of conditions of use and information, precisely because it is a nonsense word: we do not know what *noj* denotes, be it a process, an event, an individual or whatever. And this fact is crucial, for it is often assumed that the questions of what noj is and what conditions must obtain for *noj* to be used, yield the same answers. The first question is *prima facie* an ontological one: it asks what constitutes a noj. And such questions are typically asked of natural scientists. The second question is central to the thesis. It asks what conditions underly uses of the word *noj*, a question of a more obviously semantic nature.

Thus, associated with the study of word meaning, one often comes across a semantic question and a related ontological question. It is a contention of this thesis that whether these questions yield the same or different answers is itself an open question, and one that is subject to empirical enquiry. That is, borrowing a little from Wittgenstein, we must not simply say that these questions *must* yield the same answer, rather we must *look and see*. I hope that, in some way, this thesis may be interpreted as an attempt to do just that.

1.2 Some Puzzles of Word Meaning

The sorts of puzzle which we will consider cannot possibly subsume all the interesting puzzles of word meaning. Each, however, can be taken to illustrate the general difficulty of accounting for the vagueness of words and the inherent defeasibility or non-

monotonicity involved in their uses. That said, each puzzle either illustrates a different aspect of word meaning for which some account must be given or some facet of these aspects that is suggestive of an analysis.

The first puzzle illustrates the simple fact that the uses to which a word can be put depend greatly on the *situation* of use. In some situations the word can be used, in others, it seems, it cannot. Further, depending on the situation, the same word may be used seemingly to mean different things. The second puzzle illustrates the fact that provided the situation of use is of a particular kind, words may be used to refer to all manner of objects to which they would not ordinarily refer. The third puzzle illustrates an important issue to do with vagueness. That is, not knowing whether the current situation is of the right kind can lead to equivocation and disagreement about whether or not a given word can be used in that situation. So it illustrates the fact that vagueness may be seen as a property of a word, a situation and the agent's knowledge of that situation, rather than simply as a property of the word.

Finally we come to the fourth puzzle. This puzzle illustrates the problem of deciding upon *any* conditions of use for a word. Such is the variety of circumstances under which a word can be used, even when it is used conventionally, that it might seem that outlining such conditions will be a matter for natural science and not the humble semanticist. This puzzle is interesting for more than just this, however, for, of all the puzzles, this exemplifies the kind of puzzle which has traditionally attracted most attention. Not only have philosophers discussed the issue of natural kind terms but, particularly in the psychological literature on concepts, the facts about natural kinds have been seen as an important difficult test-case on which theories can cut their teeth. With the distinction made in section 1.1 borne in mind, we will argue in Chapter 3 that these puzzles can all be assimilated to the same framework.

1.2.1 The Lion Puzzle

Let us consider the word *lion* and some of its uses. In the context of a zoo or circus one seemingly uses *lion* to mean "real lion". But imagine a different context, that of a museum or art gallery. Then, it seems, one would use the word *lion* to mean "statue of a lion" or "picture of a lion". That is, there appear to be two different senses of *lion.* If indeed this is the case, then one could imagine some context, one neutral with respect to these senses, under which *both* these senses could be used to talk of one and the same object. Importantly, however, we would expect the difference in sense to manifest itself in differences in the way that the word could be used with each sense. Consider the following as an example.

> Fred is sitting on a park bench in London. He knows that at the other end of the park there is a statue of a lion. An art student approaches him and, explaining that her still-life assignment is to sketch a lion, she asks Fred if he has seen one. Fred replies that he has and points her towards the statue. A little later an exhausted zoo-keeper appears and, explaining to Fred that a lion has escaped from the zoo, he also asks Fred if he has seen one. Fred replies that he hasn't.

Fred seems to be adopting different attitudes to the same question: *Have you seen a lion?* In one situation he replies affirmatively, in the other negatively. How are we to characterise Fred's responses? We might be tempted to say that Fred is being flexible with the truth but such a characterisation seems inappropriate: characterising one response as true, the other as false, would not help us in capturing the fact that both seem equally felicitous. An alternative might be to claim both utterances to be veridical but that "lion" is simply ambiguous: that is, Fred's utterances employ different meanings of "lion", one meaning for real lions, another for statues of lions. Quine (1960), for example, suggests that "light" is ambiguous in just this way. He claims that one "clear condition" of ambiguity is the fact that the application of an ambiguous word to certain objects can be both affirmed and denied: such terms can be "clearly true or clearly false of one and the same thing" (§27 p. 131). A raven's feather, for example, is both light (in weight) and not light (in colour). Similarly, a

particular financial institution may be both a bank (a money bank) and not a bank (a river bank). Fred's case seems similar, but whereas we may be led to claim, intuitively, that Fred's statue both is, and is not, a lion, there is an important difference in the cases of "light" and "bank". In these latter cases, the different meanings of these words are deemed independent and unrelated (at least, synchronically). In the case of Fred's uses of "lion", the two meanings are highly related and indeed it is crucial to the success of his utterances that this is so.

In a similar vein, it seems, we wouldn't be tempted to claim that "lion" is simply vague as "heap" and "bald" are considered vague. It is not that the application of the word "lion" to the statue is fuzzy; rather, there is a clear sense in which we can say of the statue that it is a lion and a clear sense in which we can say that it is not. Broadly speaking, it seems then that there are two sorts of meaning attached to "lion": a *core* meaning, to do with *real* lions, and a *peripheral* meaning, to do with other kinds of lion, which is *related* to the core meaning in a particular way. Which meaning the word carries on a particular occasion of use seems to depend on the informational requirements of the agents involved which, in turn, seem to depend on the sort of situation they find themselves in. An adequate theory of word meaning must explain the relationship between core and peripheral word meaning while doing so in a way that supports the basic intuition that the application of a word to an object crucially depends on the nature of the situation in which the word is used.

1.2.2 Nunberg's Ham Sandwich

Nunberg (1978) discusses a particular case of metonymy (see also Lakoff, 1987): a waiter employs the phrase "ham sandwich" to refer to a customer who has ordered a ham sandwich. This puzzle is similar to the Lion Puzzle though it offers a starker illustration of the fact that words may apply to all manner of objects to which they would not ordinarily apply. In this case "ham sandwich" applies to an individual (a customer) who has none of the properties of ham sandwiches which make them ham sandwiches.

Similar considerations apply in this case as in the Lion Puzzle. Claiming that "ham sandwich" is ambiguous ignores the important relations between its various uses. Additionally, since we can imagine numerous similar examples, to analyse these in terms of ambiguity would suppose a quite vast, and probably implausibly vast, lexicon. Indeed, this is very much the point which Clark (1983) suggests undermines traditional, dictionary theories of word meaning. Similarly, just as before, claiming that "ham sandwich" is vague or questioning the veridicality of the waiter's utterances also seem implausible.

It is also clear that in this case it is the situation in which the waiter finds himself that allows him to make the particular use of "ham sandwich" that he does: if he said the same thing in describing a total stranger in a launderette, it is unlikely he would convey the same meaning. What we require of a theory of word meaning is that it allows the sort of defeasibility illustrated by all these examples to be very much dependent on the nature of the situation of utterance.

1.2.3 Lakoff's Mother and Macken's Father

In this section we actually give two puzzles though they are so alike we can treat them as one. The first comes from the work of George Lakoff (Lakoff, 1987); the second from the work of Betsy Macken (Macken, forthcoming).

Lakoff is concerned to expound a theory of cognitive models and, by this means, to discount the standard interpretations offered by prototype theorists of prototype effects. His article is wide-ranging, but we will focus on one aspect, the one most relevant to this discussion, that of the concept *mother*. Lakoff suggests what, perhaps, we already suspect: the concept *mother* cannot be given by a definition. There are no simple sets of conditions which are singly necessary and jointly sufficient for the application of the concept. Instead, Lakoff suggests that *mother* is defined in terms of a number of cognitive models each of which itself receives a definition. He lists five:

1. *The Birth model:* the person giving birth is the mother.

2. *The Genetic model:* the female contributing the genetic material is the mother
3. *The Nurturance model:* the female adult who nurtures and raises a child is the mother of that child.
4. *The Marital model:* the wife of the father is the mother.
5. *The Genealogical model:* the closest female ancestor is the mother.

These different definitions serve to indicate the many different kinds of use to which the word *mother* can be put. Indeed, the word can be used discriminatively so as to pick out one model as opposed to another. Lakoff offers the following example where *mother* is deemed to have the meaning given by the nurturance model.

I am not a nurturant person, so I don't think I could ever be a real mother to any child.

What is clear, then, is that kinship terms such as *mother* do not lend themselves to clear definition and that there are cases where *mother* can be used to apply to some mothers and not to others. And it is precisely this problem which Macken's analysis illustrates. Her problem is perhaps best observed in a dialogue which she recorded between a small group of six and seven year olds. Their statement of this problem and, after a brief interlude, their casual dismissal of it is so appealing we reproduce it below.

> Hasan: My father...
> Richard: You don't got no father.
> Cassandra: He got a father.
> Richard: Not his real father.
> Cassandra: He got a father that lives with him.
> Hasan: My father gonna buy me a bike.
>
> *(Macken, forthcoming)*

What the dialogue reveals is that Richard and Cassandra have different conceptions of what may and may not be called a *father.* Alternatively, they have different ideas about the kinds of situation in which the word *father* can be used. To begin with, Richard

is adamant: the word cannot be used to describe an aspect of Hasan's circumstances. However, after some discussion, Cassandra insists that the word is appropriate because there are two *kinds* of circumstance under which it is appropriate to describe someone as a father: when certain biological facts hold and/or when certain social facts hold.

Thus, the puzzle indicates several interesting facts. One is that the vagueness of words such as *father* should perhaps not be considered to be a property of the *word* as, for example, accounts in terms of fuzzy set theory assume (cf. Osherson & Smith, 1981; Zadeh, 1965), but as resulting from an ability to decide whether the current situation is of the appropriate type for the word to be used. So the content the word can convey does not have to be expressed in "fuzzy" terms, for it is not the *word* that vagueness involves. Rather it involves determining whether a given situation is of a given type. Another interesting fact revealed by this puzzle is that, associated with one and the same word may be several different sets of conditions of use, and that these different conditions of use are constitutive of the different senses the word has.

1.2.4 Putnam's Lemon

In trying to represent certain aspects of the meaning of words, it seems we want to capture certain generalisations about the properties of the objects to which those words apply. Though puzzles like the previous one may convince us of the difficulty of stating whether a given word applies to a given object, we may be tempted to treat such puzzles as special cases. Accordingly, we may consider the objective of a theory of word meaning to be the identification of those conditions which allow a word to apply to, and only to, objects which it *undeniably* describes. We might, for example, consider the conditions which allow the word "lemon" to apply to objects which most certainly are lemons and not to objects (like statues) which most certainly are not. Even though the justification for such a distinction may fall foul of our discussion of section 1.1, even with such a restriction, such conditions are notoriously difficult to isolate. That is, even restricting attention to entities which really are lemons, the arguments of Putnam (1975) testify to the difficulty in divining a set off conditions of use: though we might want to say that

lemons are oval-shape, we can easily imagine discovering one that has been squashed and, similarly, we can imagine finding sweet-tasting lemons, lemons painted red and so on. Virtually every property one might want to ascribe to the class of objects that are lemons can seemingly be defeated. That is, for virtually every property, we can imagine counterfactual circumstances under which a lemon does not possess that property. The problem then is how a theory of word meaning can capture generalisations concerning the properties of a class of objects *and* allow for defeasibility: how is it that we can infer that an object described as a lemon is oval and acidic-tasting and yet accept that some lemons can be flat and sweet?

1.2.5 Complex Nominals

In perhaps all the literatures of cognitive science which have a bearing on issues to do with words and their uses, complex nominals provide a source of often acute embarrassment. An example such as *stone lion* indicates why, for the entity referred to is not generally regarded as being both a stone thing and a lion. That is, the complex nominal is not intersective in the way that *red car* is: a red car is both a car and a red thing. Complex nominals such as *stone lion* are more similar to standard privatives such as *former senator* and *fake gun* than they are to intersective ones. Indeed, Franks (1989) has argued that *stone lion* is a functional privative. Now, the phrase *complex nominal* derives from Levi (1978) in which she discusses how the semantics of various compounds may be considered to be compositional. Her own view, which is that compounds are the surface-structure manifestations of transformations of a deep-structure representation, is not of direct interest here. Rather, her examples are. Table 1.1 lists a number of complex nominals and the categories to which Levi assigns them.

Nonpredicate NPs such as these provide difficulties for theories of semantics precisely because they are non-predicative. That is, the meaning of the complex nominal is not the same as that of the head noun with the adjective in post-copula predicative position. That is, a *stone lion* is not a *lion which is stone*, a *ham sandwich* is not a *sandwich which is ham*. Most other adjectives, however, function in the opposite way. That is, the

Expression	Category
Stone lion	Nonpredicate NP
Ham sandwich	Nonpredicate NP
Biological father	Nonpredicate NP
Adoptive father	Nonpredicate NP

Table 1.1: Some Complex Nominals (following Levi, 1978).

designation of the NP is given by the intersection of the designations of the adjective and noun. Compounds such as *red car*, *wet dog* and *tired writer* are generally of this sort. And the fact of their intersective nature is typified by the fact that such compounds are predicative. A *tired man* is, indeed, a *man who is tired*, and so on.

Now, in the case of *stone lion* we may take the locution *lion which is stone* as only partly and inexactly specifying the meaning of the complex nominal. The fact that such paraphrases often seem to require some additional component or relation such as *statue which is made of* in the case of *stone lion* has led some linguists (cf. Kay & Zimmer, 1976) to suppose that the interpretation of compounds such as these are potentially infinite. That is, their interpretation depends on an infinite number of relations which mediate between the components of the compound.

What we have shown in the previous sections, however, suggests that the problems that are posed by nonpredicate NPs are, in some sense, subsumed by the puzzles of *single* word use. If we consider the case of the Lion Puzzle and the related nonpredicate NP, *stone lion*, we can see that *lion* appears to have two senses. One can apply to statues of lions, one applies only to real lions. The point being that our analysis of the nonpredicate NP *stone lion* will depend on which sense we assign to the head noun *lion*. If, for example, we assign it the sense "statue of a lion" then there is now no problem in analysing the nonpredicate NP. Indeed, its meaning can be seen to result from the usual composition of the meaning of adjective and noun. That is, if this is the sense of *lion*, then the NP is intersective. For *stone lion* translated in this way simply means *stone statue of a lion*. So, it is in this sense that we intend the puzzles we considered earlier in this chapter to subsume the related puzzles provided by the corresponding

nonpredicate NPs.

1.3 Concepts and Word Meaning

Having distinguished, in section 1.1, the question of which conditions must hold for a word *X* to be used and the question of what an X is, we are now free to indicate what the puzzles entail. Before we do, let us re-iterate our view of the uses of words, their conditions of use and the informational content which they may possess.

Prima facie, the puzzles of the lion, the ham sandwich and the mother/father seem to indicate that the same word can carry different information on different occasions of use. That is, associated with each word and the context in which it is used is a set of conditions. It is the obtaining of these conditions, we assume, that licences the use of the word. And, moreover, it is the obtaining of these conditions that invest these uses of words with informational content. So, we will assume that a competent speaker of the language will generally use words only if the conditions of their use are assumed to hold in the context of use. And, in general, this will be assumed unless there is evidence to the contrary. That the speaker's assumption may be wrong can lead to cases of misinformation. But, that the speaker's assumption may be correct, that is, in cases where these conditions do hold, their obtaining allows the use of the word to carry information to that effect. That is, a competent hearer of the language, realising that the word may be used only in certain circumstances, takes the speaker's locution as evidence that these circumstances hold. The hearer, then, assumes the use is informational unless there is evidence to the contrary.

Now, the puzzles suggest that associated with each word there may be many sets of conditions, each set relevant to particular kinds of circumstance and corresponding to different conditions of use. *Lion*, in the circumstances where an exhausted and increasingly frantic zoo-keeper is involved, seems to have the informational content normally associated with the locution *real lion*. And *lion*, in the circumstances where an art-student, a museum curator or a parks attendant is involved, seems to have informational

contents normally associated with locutions such as *statue of a lion*, *picture of a lion*, etc. And the case is similar in the other puzzles. *Ham sandwich* seems to have different informational content associated with its different uses; *father*, too, seems to have different contents associated with its uses, a fact of which Hasan, no doubt, is painfully aware.

Now there are many interesting questions to do with this difference in content, but the primary one from the point of view of psychology concerns the relation between these contents and the contents of mental representations. To attempt to explicate this relation, we will consider two related topics in cognitive psychology. One is concerned with "dictionary" theories of mind, the other to do with theories of concepts.

In trying to explain how people both understand and process language, theorists have found the analogy with dictionaries appealing. As Clark's work shows (Clark, 1983; Clark & Gerrig, 1983; Clark, 1989) the analogy has often been taken as much more. That is, the assumption behind models of human parsing has been that, at some point during parsing, a "mental lexicon" is accessed. Entries in the mental lexicon are compared with the input, segmented into words, in terms of orthographical or phonological properties and, that done, the "sense" associated with the word is "read off" from the dictionary entry. In cases where there are multiple senses for a given word, the appropriate sense is selected from a list of possible senses. Now, we will have much more to say about this in Chapter 3, but for now it is sufficient to note some of the characteristics of such dictionary theories.

Firstly, they associate with words some descriptive content. This content, it is assumed, guides linguistic behaviour, that is it guides, or determines, the appropriateness or otherwise of any use of a given word. Secondly, such theories presume that this content is mentally represented in a certain way. Namely, the assumption is that any sense of a given word is permanently represented, just as in a dictionary, and that parsing constitutes the selection from such senses. So, there are two claims: one is to do with the descriptive content that can be seen to be implicated in linguistic behaviour; the other is to do with how such content is represented.

The second sort of theory in which we will be interested concerns the psychology of concepts. From observations concerning the discriminability of behaviour, the conjecture has been made that guiding such discriminations is a descriptive content, labelled a concept. For example, a child sorting a variety of objects into two heaps, lemons and non-lemons, is taken as evidence that the child has the *concept* lemon. Just as with the dictionary theories, there are two claims often associated with theories of concepts.

One is the assumption that the criteria used in sorting lemons from non-lemons are constitutive of the concept. That is, such criteria reveal the descriptive content associated with the concept. And it is also assumed that such descriptive content guides behaviour, that is, it determines the appropriateness or otherwise of any use of a given concept. Secondly, there is also a claim about mental representation. That is, it is assumed that these concepts are permanently represented. That is, the assumption has been that semantic memory is comprised of a finite number of units: each unit comprises a certain descriptive content. Determining the appropriateness of certain behaviours is constituted by the process of finding the appropriate unit, a unit whose descriptive content is satisfied by the current circumstances. These units, then, are concepts. Now, this last process, that of finding the right concept for the circumstances, has always been recognised to be hard. But, nonetheless, the picture which most theories of concepts paint is the one we have described. In many respects, it is a picture which bears a striking similarity to the dictionary theories we described above. Indeed, as we shall see, the prevailing view is that the contents of the mental lexicon are none other than concepts. However, for the moment, let us return to the puzzles.

The first three puzzles all seem to involve the *partitioning* of informational content into the information associated with unconventional uses and that associated with conventional uses. That is, the information associated with X where something isn't really an X and the information associated with X when something really is an X. So, for example, the statue of a lion in the Lion Puzzle is not really a lion, though the *word lion* may apply to it. Similarly, the customer ordering the ham sandwich isn't really a ham sandwich, though the words *ham sandwich* can be used to refer to that customer. And likewise in the case of Hasan's father. Richard is adamant that a "live-in" father

is not a *real* father, but nonetheless he can concede that the word *father* may apply to such individuals. Now, although we are not so much concerned with the conditions under which something is an X, we are concerned with all cases in which the word *X* carries informational content. So, during the course of this thesis, we will be concerned with two issues. We will be concerned to distinguish cases of the same word carrying different information. And we will be concerned to determine the ramifications of these cases for claims made both by traditional dictionary theories and theories of concepts.

The last puzzle is most illuminating given our stated aim of exploring the literatures on concepts and word meaning, for, the puzzle of Putnam's Lemon is worrying for anyone concerned with questions of informational content, descriptive content and conditions of use. As our earlier discussion indicated, arguments have been advanced to the effect that natural kind terms are simply *non*-descriptional. Such a conclusion is indeed disturbing for one whose aim is an explication of the conditions under which such words may be *used* and of the informational content associated with such uses. Are we, then, to conclude that language users are unaware of the conditions under which the word can be used?

There are several possible answers. One, interestingly enough, comes from the psychological literature on concepts, in particular, from prototype theory. The work of Smith & Osherson, for example, has been to explicate concepts in terms of descriptive content which is based on the prototype for that concept. So, the claim is made (Smith et. al., 1988) that associated with the concept for *apple*, say, is a certain cluster of descriptions. Further, the claim is advanced that these descriptions are implicated in the *composition* of concepts, such as *red* and *apple*, for example, in forming the complex concept *red apple*. Now were this a claim simply about the conditions of application of *concepts*, we would be none the wiser concerning the informational content associated with uses of natural kind terms. However, the literature on concepts is much more closely tied to the literature on word meaning. Johnson-Laird (1987), for instance, proposes a theory of word meaning in which the senses of words are represented in a mental lexicon, the entries of this lexicon being *prototypes*.

We will have more to say of this correspondence between word meaning and concepts later. Indeed, Chapters 5 and 6 should be seen as an attempt to draw out this correspondence more clearly. The striking similarity between positions adopted in the philosophical literature on word meaning and the psychological literature on concepts is suggestive that a comparison of these approaches may be mutually illuminating. Indeed, in Chapter 7, we take seriously the identification Johnson-Laird suggests, and consider the ramifications for prototype theory of considering Sense Generation as a theory of concepts. At any rate, returning to the puzzle of Putnam's Lemon, we can say that the descriptive content we associate with uses of natural kind terms will depend on how we treat the psychological literature on concepts. And this we do in Chapter 6.

Before we leave this section, we must draw a further distinction, one which has important ramifications for the representational claims associated with theories of concepts and of word meaning. The distinction concerns two views of concepts, two views of the senses of words.

Concepts and senses have been seen in two lights: firstly, as effective classifications of various behaviours and, secondly, as mental represented objects. So, for example, dictionary theories have been taken not simply as descriptions of conditions under which a word may be used, but as claims about how knowledge of such conditions is mentally represented. For instance, the debate between decompositional theories of the mental lexicon (cf. Johnson-Laird, 1987) and meaning-postulate theories of the mental lexicon (Fodor et. al., 1980) do not differ in their claims about *words*. They both claim, for example, that *bachelor* is appropriate for single, adult males. Where they differ, however, is how this knowledge is to be encoded. Decompositional theories, for example, hold that there are discrete lexical entries which contain a decomposition of the sense of a word in more primitive terms. That is, the analogy with a dictionary is very strong indeed. Meaning postulate theories, however, view a lexical entry not as decomposed into more primitive terms but as being related via meaning postulates to other entries. So, although these approaches are *logically* equivalent, it is thought that they make different claims about how senses are mentally represented.

When we come to discuss dictionary theories and concepts, we must be clear about the relation between these two claims. In particular, we start from observations about word use. We will, in Chapter 3, formulate a theory of how it is that the same word may have numerous different senses. So, our theory will not *prima facie* be about representations. Rather, it will be concerned with the appropriate *classification* of word uses and the information they can carry. Armed with this classification, we will leave open the question of how this classification is represented. Rather, the theory is intended as specifying constraints on what such representations must encode, what they must compute. In particular, the theory should be seen as specifying, in a certain way, the nature of the relationship between words and their informational contents. How this relation is to be interpreted in terms of representation or computation is another matter.

Let us be more abstract. Consider our old friend *noj*. And let us suppose that this word has a number of different informational contents, $ic_1 \ldots ic_n$. There are many ways we can describe the relationship between these contents and the word *noj*. We could describe it in terms of a single relation between *noj* and $ic_1 \ldots ic_n$. We could consider the relationship in terms of some two-place function mapping *noj* and some entity to a single informational content. We could consider the relation in terms of a function composition. For example, we could consider the composition of a function mapping *noj* to x and a parameterised function mapping x and a context to an informational content. We could consider the relation in terms of two functions, three functions, and so on. In fact, given such a relation, there are an arbitrary number of ways of describing it. What is crucially important for our purposes is that the functional description we employ makes sense given our interest in the literature of cognitive science. Indeed, we take it that the various disciplines of cognitive science place constraints on what counts as an appropriate functional description. That is, we will not be talking in great detail about mental representation, because for each functional characterisation we employ, there is a potentially infinite number of different ways that function could be computed. So, even equipped with an appropriate functional characterisation of the relation between a word and its informational contents, such a characterisation is consistent with many different claims concerning mental representation. That is, one cannot advance one

particular claim on the basis of such a characterisation. However, though it is logically possible to derive many different functional characterisations for the relation between a word and its informational contents, we place the constraint that such a characterisation must be appropriate to the various disciplines of cognitive science. That is, it must be philosophically appropriate, psychologically appropriate, formally appropriate, and so on.

Now, when we consider standard theories of sense, we can see two possibilities for error, both in their representational claims and in their functional claims. Firstly, with respect to the functional claims there is the possibility that the functional description they employ is simply not appropriate. So, for example, the relation between a single word and its uses is standardly considered in terms of a single function. This is what is assumed by theories which are committed to the view that all the uses of a word, or even simply all the conventional uses of a word, must have something in common. And the meaning must specify what these commonalities are. Now while such a functional description *may* be appropriate, it is by no means certain. Indeed, we shall argue in this thesis that it is not appropriate.

So one source of the inadequacy of a theory of word meaning is whether or not the functional description employed is appropriate. Another possibility for error arises when this functional description, whether appropriate or not, is taken as directly revealing of mental representation. So, for example, a meaning postulate theory of word meaning may be, according to some views, an appropriate description of certain aspects of word meaning. However, if it is seen as directly revealing of the way in which that aspect arises from mental functioning then another source of error may arise. That is, the facts of the matter about representation, the computational primitives of neural architecture, for example, may be at odds with the way such a functional description is formulated. Indeed, this is hardly surprising since it is a standard result of the theory of computation that a given function may be computed in arbitrarily many ways. So, when a particular functional description is taken as a statement of how that function is computed, the possibility of error will arise. Consequently, we will not claim that the theory we offer later is committed to any particular facts about *how* this theory might be implemented.

Similarly, although we hold to the view that certain aspects of the theory *are* mentally represented, we do not make claims concerning *how* they are represented.

To recap, then, our position is one which treats theories of concepts and sense as similar in certain essential respects. They are similar inasmuch as the functional descriptions they attempt to characterise are similar. Their precise relation we only guess at later. Our position, however, is that the functional descriptions they employ are not appropriate given considerations in the literature and, therefore, we attempt to provide a more appropriate functional description. Concerning representational claims, our position is that the nature of representation is logically independent of the nature of the function computed: that one cannot advance from a position on the latter directly to a position on the former. Inasmuch as representational claims have efficacy, they do so only to the extent that they sit with the facts about representation, the facts about the computational primitives of a neural architecture. As we do not touch on this issue, we make only limited representational claims. Our claim is simply that the theory we offer later amounts to a better functional description of issues to do with word meanings, concepts and senses than those offered previously.

The question to which we now turn is that of what these senses and concepts are senses and concepts of. To answer this question, we turn to a discussion of Situation Theory.

1.4 Word Meanings as Informational Constraints

Barwise & Perry (1983), in *Situations & Attitudes*, spell out what they call the Relation Theory of Meaning. Central to this theory is a view of meaning relations or constraints in which these constraints play a dual role: that is, they are seen both as meaning relations and as carriers of information. These two aspects of constraints are also suggestive of the wider concerns of Barwise & Perry's theory.

For Barwise & Perry, the study of meaning is intimately wedded to a study of information flow. That is, it is concerned with the conditions under which information can

flow and the extraction of this information from situations and by agents. By locating various kinds of meaning within an overall picture of information, by concentrating on the information that agents can extract and on the exploitation of aspects of situations by agents in conveying information, Barwise & Perry offer a less dogmatic approach to meaning. By this we mean to distinguish this theory from traditional theories of meaning. Though we will not consider these here, we will focus later (Chapters 5 and 6) on the approaches of Kripke and Putnam which assume a traditional analytic conception of meaning.

Barwise & Perry offer an incomplete taxonomy of constraints. There are two broad classes of constraint: conditional and unconditional. We can further subdivide these classes into nomic, necessary and conventional constraints. Before explaining what these last are, we will consider the distinction between conditional and unconditional constraints. But before we may do this, we will have to consider some of the building blocks of Situation Theory.

In Situation Theory we can talk of situations in terms of the facts which they support. We model facts in terms of infons and we write that a situation s supports some infon σ as follows:

$$s \models \sigma$$

Facts are composed of some relation and some number of individuals which are appropriate for the argument roles of that relation. We write that a fact σ has a relation R and arguments $a_0 \ldots a_n$ as follows:

$$\sigma = \ll R, a_0, \ldots, a_n; + \gg$$

We indicate that such arguments do stand in the relation R by assigning the infon a polarity, $+$; a polarity, $-$, would indicate the opposite. Now, facts may or may not

be supported by particular situations and we write that the fact σ is supported by a situation s, as follows.

$$s \models \sigma$$

Now, by these means we can also talk of situation types. That is, we can talk of the type of situation such that situations of that type support σ. We write the situation type, S, as follows:

$$S = [\dot{s} \mid \dot{s} \models \sigma]$$

Constraints can also be viewed as facts. In this case, there is a relation "involves", written $\Rightarrow$, which has two argument roles associated with it, roles for which situation types are appropriate. So, if a constraint is supported, we can write this as follows

$$s \models \ll \Rightarrow, S_1, S_2; + \gg$$

where S_1 and S_2 are both situation types. This means that, firstly, a situation s supports a fact and this fact is a constraint. And, secondly, that the constraint holds between two situation types S_1 and S_2 such that provided there is a situation of type S_1 then there is one of type S_2.

Now we are in a position to say what an unconditional constraint amounts to. Unconditional constraints are ones whose existence is independent of any particular situation. That is, there are no conditions on s for the constraint to be a fact. Alternatively, all situations, regardless of their type, will make such a constraint a fact. That is, such constraints are deemed to be universal, components of all situations.

Now conditional constraints are of quite a different nature. For a conditional constraint to be a fact, it must not only be supported by some situation s, but that situation

must be of a certain type. Barwise (1985) offers an alternative way of writing this. Conditional constraints could be written

$$S_1 \Rightarrow S_2 \mid B$$

where S_1, S_2 and B are all situation types. B refers to the background situation type, the type that must be realised if the constraint is to be supported. Provided this situation type is realised, then the involves relation holds between the specified situation types. But what this amounts to is the claim that unless certain conditions, background conditions defining the type B, hold in the current situation then within that situation the constraint does not hold. That is, within such situations there is no meaning or informational relation between types S_1 and S_2.

We can picture conditional constraints as indicated in Figure 1.1. That is, we can view the relativity of conditional constraints as holding to a situation, b. This, Barwise tells us, offers a better viewpoint on the seemingly mysterious background conditions defining the situation-type B. Thus, any situation which would previously be of the type B, defined by these background conditions, will now be seen as a smaller situation of b. A conditional constraint relativised with respect to b in this way, can then be seen as a fact which is a component of all situations which are included in b. In all situations smaller than or equal to b, therefore, the informational relationship between S_1 and S_2 holds; in other situations, the relationship does not hold. One can see, then, that a conditional constraint can be thought of as a "local" version of an unconditional constraint: provided an agent remains within certain situations, defined in terms of b, the informational relationship which is the constraint will appear as an entailment, as a necessary relation, as unconditional. The theorist, having privileged access, can see that the constraint is really of a conditional nature but, of course, the theorist herself is limited: constraints that appear unconditional may, in fact, simply be conditional constraints. The appearance of unconditionality, then, is one that can arise from the inability to consider all relevant situations.

There are a number of other sorts of constraint which we will just briefly mention:

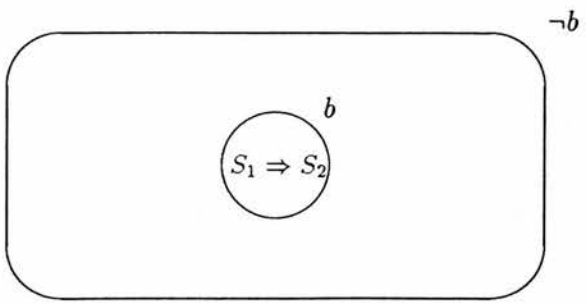

Figure 1.1: A Picture of Conditional Constraints.

nomic, necessary and conventional. Nomic constraints are ones which reflect natural laws. The relation between smoke and fire, for example, reflects natural laws. Necessary constraints are exemplified by relations between, for example, kissing and touching. It is deemed necessary that every kissing is accompanied by a touching. Conventional constraints reflect relations that can be flouted. For example, it is a convention of language that one says what one believes, but one can be dishonest, thus breaking the convention.

Before turning away from constraints, we should note that in postulating the existence of constraints, there is at least a dual commitment. One commitment is to the logical nature of the relation specified by the constraint. For example, if we propose that there is a conditional constraint relating x to y, then, duly, the nature of this relationship should be conditional. Similarly, we make a commitment to the relata of the constraint. That is, we make a commitment that these are the terms of the constraint to which agents may become attuned. If, for instance, no agent behaved in a way consistent with this relation between x and y, but did so consistent with a relation between $x \wedge z$ and y, then the latter might be seen as revealing the true nature of the constraint. A related commitment is an ontological one. Constraints are facts just as the fact that I, at the

time of writing, am writing this thesis. In the ontological commitment we make, there is little difference. Consequently, we must be careful when postulating the existence of constraints. We must be careful, for example, in postulating many constraints as opposed to just one, as we ought to be careful in postulating the existence of an object corresponding to every possible vantage point we may have on a single object. To extend the analogy, suppose that in looking around me I judge there to be six chairs in the room. Now, it is presumably not beyond the realm of possibility that there are seven chairs in the room but that there is some mysterious optical illusion which only allows me to detect six. Now, were I to claim that there are six chairs in the room, I must accept the possibility of error. What looks like six chairs might, in fact, be seven. And, in postulating constraints, there is a similar danger. What may look like one constraint might, in fact, be two and what may look like two might, in fact, be one. Since, in postulating constraints we make claims about the structure of the world, we must, consequently, exercise a certain caution.

So much for the detail of constraints. Stepping back a little, we can see that, in situation semantics, there is an entwining of meaning and information and, we may suggest, this entwining offers a potentially richer view of semantics. So, for example, Barwise (1989) tells us that "such a theory will have more to do with the ordinary use of the word *meaning* than with the use that identifies meaning with linguistic or conventional meaning" (p. 50). It is the possibility of this richer view of semantics which, given the concerns we expressed in the first section, is appealing. However, we can focus on a number of aspects.

Barwise (1989) distinguishes several types of meaning. Firstly, there is what he himself chooses to call meaning, namely *sentence meaning.* Consider the sentence *I admire Vygotsky.* If I utter such a sentence then I express a proposition which is undoubtedly true. If someone else utters the same sentence, then the proposition they express will be different, indeed it may turn out to be false. Firstly, the situation-semantical enterprise is one of turning away from analysing propositions solely in terms of their truth value. Instead, it proposes to treat more seriously the "subject matter" of propositions. For example, my utterance of *I admire Vygotsky* does not simply express a proposition which

is true, it expresses a proposition about myself and about my disposition towards the works of Vygotsky. This notion of subject matter marks, then, one of the departures of situation semantics from more traditional approaches. Another, concerns the other types of meaning Barwise discusses.

Sentence meaning, then, is that which is common to all utterances of a given sentence and it is necessarily of rather an abstract nature. *Content*, however, is rather different. The particular proposition I express by my utterance of *I admire Vygotsky* is, in some sense, the meaning not of a sentence but of a particular utterance. It is the meaning of the particular event where I utter the sentence. This type of meaning is what Barwise calls *content.* The third type of meaning which Barwise discusses is that of intended meaning, what a given speaker of a sentence intends to convey. This, Barwise calls *user meaning.* If I am not a competent user of English, for example, then I may utter *I admire Vygotsky* fully intending to convey just the opposite. This, then, would be a case where user meaning and content may diverge.

In discussing the puzzles of section 1.2, we need to ask which type of meaning we need to examine. Consider Fred, the statue of the lion and the sentence *this is a lion.* When Fred utters this sentence while pointing to the lion, it seems Fred may adopt two different attitudes to his utterance. Alternatively, he may intend two different meanings for the same sentence. But, it would not be right for us to simply analyse the puzzle in terms of what Fred intends to mean, for, as the puzzle indicates, Fred's interlocutors are capable of grasping the *same* meanings. So the notion of meaning we need is closer to Barwise's notion of content. Let us look at this in more detail.

An utterance of the sentence *you are likely to fall asleep* expresses some proposition. The content of the proposition can be thought of as being determined by two general factors: the meaning of the sentence and the circumstances of utterance. These latter will, for instance, determine the referent of the indexical *you.* So we can write the propositional content of the sentence as $C(S,c)$ where C is the content, S the sentence and c the circumstances of utterance. So, we have a picture of how the content of an utterance can vary not only with the sentence uttered but also with the context of utterance.

For sentences such as *Leo is a lion*, where Leo is the proper name of the statue of a lion near to where Fred is sitting, the picture is a little more difficult. Whereas it is possible to suggest that the content of the sentence, $C(S, c)$, differs on the occasions of Fred's utterances because the contexts of use, c, differ, this compels us to the view that words such as *lion* are like other indexical expressions. It may well be that the context sensitivity of expressions in the puzzles can be formally distinguished from that of indexicals. Accordingly, though we will relate our discussion of the senses of expressions to this notion of the content associated with utterances, we will not be committed to the identification of the properties of common nouns with those of other indexical expressions. The prospect of such an identification would take us outside the scope of this thesis.

We will finish this section by commenting on the background conditions with respect to which we may construe conditional constraints as being relativised. Conditional constraints, as we saw in Figure 1.1, can be conceived as facts which hold only in certain situations, situations which are smaller than or equal to, say, b. Now, the conditions that were previously employed in defining situation-types, B, can be better understood. For, previously, if someone were to ask the nature of this background situation-type, that is, ask the nature of these background conditions, we would encounter problems. Firstly, it is not clear that an answer would ever be forthcoming and, secondly, it was not clear that it wasn't a fair question to ask. I now wish to suggest, in defence of our later use of conditional constraints, that such a question is better directed at a natural scientist than it is at a semanticist. This is not due to some inward malevolence towards natural scientists, but rather that such a question is precisely the sort of question to which natural science begs answers.

Asking the nature of background conditions is equivalent to asking what conditions need be satisfied by the current situation in order for this constraint to be satisfied. That is, it is to ask why such and such an informational relation holds in a give situation. Casting constraints as facts alters this drastically. For, now, to ask why an informational relation holds in a situation is to ask why one particular fact holds in a situation. And, presumably, the only possible answer will be couched so as to explain why this fact

depends for its existence on various other facts.

Consider the fact of the chair on which I am sitting being a Chippendale. There are presumably many other facts on which the existence of *this* fact depends. Many of these will be facts concerning who made the chair, its style, the period in which it was made, and so on. But there are many other facts on which the fact of my chair being a Chippendale depends. Facts concerning the tools with which it was made, facts concerning the ability of furniture makers to communicate with one another, and facts concerning the ability to identify and maintain particular styles. Indeed, were we to articulate *all* the facts which matter as far as a chair being a Chippendale is concerned, we would have produced a complete science of furniture making and much else besides.

Now, it is similar, I am suggesting, for facts that happen to be conditional constraints which relate worldly conditions to types of uses of words. Clearly, some of the conditions on which their existence depends will be of considerable relevance to a semantic enterprise. I would not suggest otherwise. What seems likely, however, is that an articulation of the background conditions with respect to which a conditional constraint is relativised is not a goal that should be expected of a semantic theory. It is enough that a semantic theory should correctly characterise the informational relationship between, for example, words and their meanings. It should not also have to articulate all those facts on which the existence of this relationship depends.

1.5 Concepts and Information

The view we have sketched, namely that of viewing concepts as expressing attunement to informational constraints, raises the interesting question of in what sense concepts can be said to be informational. For, if concepts are themselves informational then how are cases of misinformation to be explained?

There are many cases where the mental states of individuals do not carry information about a situation with which they are linked. The case of the Ames room provides a

classic illustration. The Ames room is constructed so that, viewed from a particular vantage point outside the room, it looks like a conventionally arranged and constructed room. However, there is a deception in this. For, if one enters the room it soon becomes apparent that the construction is distinctively non-conventional: that the walls are not parallel and nor are the floor and ceiling. However, the construction is such that these *look* parallel when viewed from this particular viewpoint. Now, quite clearly, someone viewing the Ames room from this position can entertain various conjectures about it. One might be that the ceiling and floor do indeed lie in parallel planes. That is, someone's concept of what it is for two planes to be parallel may well be satisfied. Only on further inspection, will it seem otherwise.

Now there are several observations here which have relevance for our claim that concepts are informational. One is that the concept of "parallel planes" guides certain behaviour. Behaviour involving reaching to the floor, for example, may well be guided by the belief that this is parallel to the ceiling and consequently perpendicular to the bodily axis. In this sense, the satisfaction of the concept appears to inform the agent's actions. It also informs the agent's thoughts: thoughts concerning other facts about the room such as the distance the walls are apart from one another.

In an important sense, however, this aspect of the concept of "parallel planes" is not, and cannot be, informational. For, although concepts may guide behaviour it is clear that, in the cases of the sorts of illusion we have considered, such behaviour is not based on the actual way of the world. It is based on the way the world is viewed and, in cases where the view provides an incorrect view of the way of the world, such views do not provide information. They do not indicate that things are thus and so, they merely purport to so indicate. In this sense, then, concepts should not be seen as informational: the mere accessing or use of a concept does not guarantee its satisfaction.

But such a conclusion is not incompatible with the view that there are informational relationships to which concepts express attunement. So, for example, in order to have the concept *parallel planes*, there must be some constraints which, say, govern the behaviour of bodies that lie in parallel planes. Without these it would be difficult to imagine there

being a corresponding concept. So, when we say a concept is informational, we mean merely this: there is some constraint or set of constraints to which such a concept expresses attunement. It is not to say that whenever a concept is accessed or used the constraints to which it expresses attunement are part of the current situation. For to say that, would disallow the possibility of concepts supplying misinformation.

1.6 The Argument: a Preview

In this chapter we have established a framework for discussing issues of word meaning and concepts, a framework that distinguishes the ontological question of under what conditions can something said to be an X from the semantical question of under what conditions can the word *X* be used. We have presented a number of puzzles whose solution we require and have related these to the more notorious category of complex nominals. Finally, since part of our approach is to offer a formal account of a theory of word meaning, we have expanded on the formal framework of Situation Theory. These, then, are the tools we will need to begin.

In Chapter 2, we will describe a number of options that, at least *prima facie*, are open for a theory of word meaning to adopt. The first we consider are choices in how we explicate meaning. So, for instance, we will consider whether meaning relations are best seen in terms of situation theory's conditional or unconditional constraints; we will consider whether the meaning of unambiguous words is to be explicated in terms of one or more such meaning relations; and we will consider whether these meaning relations can be conceived of as being descriptional. The remaining choices in the Chapter relate to the mental representation of word meaning. Here, we will relate the choices for both the number and content of mental representations of the meanings of unambiguous words. We will also comment on the relations between these choices. Before, finishing this chapter, however, we will consider a few remaining issues as, for example, the relation between the language in which we might describe mental representation and the facts of representation. By these means, many of the possibilities for a theory of word meaning will have been enumerated.

In Chapter 3, we spell out a type of theory of word meaning, Sense Generation. We discuss, in some detail, Herb Clark's related theory of Sense Creation, and indicate the differences. We expand on his account of Sense Selection, distinguishing two types: weak and strong sense selection. The deficits of sense selection accounts lead to our characterisation of Sense Generation, where we expand on the notions employed in the theory: senses, words and lexons.

In Chapter 4, we spell out a particular theory which falls under the rubric of Sense Generation: this is the Relational View. The first part of this chapter details the motivation for this view, much of which comes from a consideration of the psychological literature on concepts. Having outlined these, we turn to a more formal description of the relational view in terms of situation theory and, where necessary, with reference to a prolog implementation. The view is initially described in terms of an example of natural kinds. Following this, however, we return to the puzzles of Chapter 1 and offer an account of their solution. At the end of this chapter, we begin a discussion of meaning and, in particular, analyticity, that will occupy us, in essence, for the following couple of chapters.

We start, in Chapter 5, by considering an influential argument that certain words, proper names, are not descriptional in the way that we have assumed for common nouns. Such a conclusion, if correct, will have significant ramifications for the general application of our theory. We begin by examining some description theories of proper names and suggest that one of the important motivations behind these theories was the possibility of giving some account of cognitive significance. We then turn to Kripke's argument in *Naming and Necessity* which stands to refute these theories. Despite the apparent validity of Kripke's argument, it results in the divorce of philosophical approaches to word meaning from the notion of cognitive significance. Rather than take this option, we suggest that Kripke's argument, while seeming to be valid, is best construed as a *reductio.* It should be seen as suggesting that the conception of analyticity which it presumes is in need of revision. Accordingly, we propose an alternative conception of analyticity that which is suggested by Sense Generation.

In Chapter 6, we discuss the related psychological literature on concepts, with particular respect to the treatment of natural kind terms. We begin by considering the traditional classical view of concepts and some of its failings, and by relating such a view to the Frege-Russell position on proper names. Following this, our attention turns to Prototype Theory where, again, its more obvious failings and its relation to the Searle-Strawson position on proper names are examined. We then briefly consider hybrid views, those which assume concepts to be comprised of one component amenable to a classical analysis, one amenable to a prototype analysis. We then return to the argument of Kripke and consider the related arguments of Putnam. Our suggestion is that their conclusion that natural kind terms are non-descriptional remains unrefuted by these psychological theories of concepts. However, by accepting the position of the previous chapter, that the Kripke-Putnam position should be viewed as a *reductio*, we suggest that these psychological views of concepts are immune to the Kripkean conclusion. We briefly spell out a Sense Generation account of natural kind terms and specify how it circumvents Kripke's arguments. Finally, we hint at a possible reconciliation between the psychological treatment of concepts and philosophical views of word meaning.

Our last major chapter, Chapter 7, is concerned with offering a re-interpretation of one psychological theory of concepts, prototype theory, from the perspective of the relational view of Chapter 4. We begin by offering both an exposition of prototype theory and expositions of two models of prototype concepts. This achieved, we offer a classification of prototype theory in terms of the choices we outlined in Chapter 2. Having developed a clear conception of prototype theory, we expound on some of its difficulties. We briefly re-iterate the relational view, though we now call it the Family of Constraints view and this establishes the position from which we can attempt the main focus of Chapter 7, a detailed comparison of prototype theory with the Family of Constraints view. Our suggestion is that this comparison suggests greater psychological plausibility for the Family of Constraints view.

In the last chapter, Chapter 8, we begin by recapping on the arguments of the thesis, its results and its implications. We conclude with a discussion of some future research

which it is hoped the work we have related in this thesis will lead to.

Chapter 2

The Choices

In this chapter we start moving towards a theoretically coherent view of word meaning. We are thus naturally led to consider a number of choices. The decisions we make in this chapter will determine the nature of our theory of word meaning. In the next chapter we will argue against various options and settle on one, Sense Generation. Here we concentrate more or less on simply what these choices are.

2.1 Meaning

As we saw in section 1.4, within Situation Theory there are a number of ways of treating meaning and information. In particular, we have a choice as to whether we analyse the puzzles of Chapter 1 in terms of conditional or unconditional constraints. However, as we shall see, this is not the only choice facing our analysis of the meaning relations underlying such puzzles.

2.1.1 Conditional or Unconditional

The choice of whether conditional or unconditional constraints are taken as meaning relations crucially depends on how we treat the issues of defeasibility and non-monotonicity.

Standardly, meaning relations are taken to be indefeasible, that is, they observe the property of monotonicity: in all contexts of use the relation itself, rather than our knowledge of it, cannot be overturned. Classical entailment relations are one sort of relation which is indefeasible and, indeed, this fact can be used in describing the property of monotonicity, as follows.

$$\frac{\Delta \models \phi}{\Delta, \psi \models \phi}$$

Thus, if in all models where Δ is true, ϕ is also true, then Δ can be said to *semantically entail* ϕ. Consequently, if any other proposition ψ is also true, then Δ and ψ is said to semantically entail ϕ. The only condition we need to determine is whether or not all models satisfy both Δ and ϕ. We need not concern ourselves with the relation between ϕ and ψ.

The assumption that meaning relations are semantical entailments is a commonplace, even though it may be claimed merely as a working assumption. And this has a bearing on various troublesome facts concerning natural language. One is the prevalence of defaults. So, the fact that birds typically fly might tempt us to suggest that *bird* means *flying animal.* Now, were this to be the case, construing meaning in terms of entailments would render it impossible for *bird* to mean anything other than *flying animal.* That is, it would be impossible to say that the designation of *bird* included things which were not flying animals. To say that *emus* are birds, for example, would amount to a contradiction. The clear fact of the error of this conclusion can be taken to imply either one of two things: that meaning relations are not entailments and they do not have the property of monotonicity; or that *bird* does not *mean* flying animal.

Consequently, if meaning relations are deemed to have the property of monotonicity, that is, they cannot be defeated, then any explication of meaning must be in terms of necessary and sufficient conditions. Indeed, it is often argued that the possibility that a claim may be defeated, renders that claim to be about something other than meaning. So, for example, the arguments of Putnam and Kripke are both of this character. They assume, for instance, that since it is not a *necessary* fact that lemons are yellow, such a

fact cannot be part of the meaning of *lemon.*

Despite some of the apparent problems of this construal of meaning relations in terms of necessary and sufficient conditions, it cannot easily be rejected. The origins of such a construal may be traced back to, among others, Kant, Leibniz and Hume, and the notion of *analyticity.* We will discuss this more in section 5.3. For now, though, we can suggest the following: the conception that meaning relations are entailments disallows the possibility that meaning relations may be overturned or defeated. If a claim concerning Xs can be defeated, then it cannot be a claim based on the meaning of *X.*

Though we will discuss later some of the analytic conceptions of meaning, we will briefly mention an alternative based on an interpretation of Wittgenstein's notion of Criteria-relations.

Baker (1974) offers us several properties of criterial- or C-relations or, alternatively, the relations of C-support. Firstly, criteria contribute to meaning: if x C-supports y, then the meaning of y is partly determined by x. Through being evidence that makes y certain, x partly determines the content of the meaning of y. Such meaning relations are given as part of the "grammar" of our knowledge of language and the world, and as such the connection between the antecedent and the consequent is *a priori.* If we do take such C-relations to partly constitute the meaning of a sentence or word, then we arrive at the conclusion that a change in C-relations will produce a change in meaning.

Secondly, the C-relation is distinct from entailment. Wittgenstein consistently contrasts criteria with necessary and sufficient conditions — it is "weaker", in that C-support is logically both positively and negatively inconclusive. This is related to the circumstance-dependence of C-support: whether particular evidence functions as a criterion for a particular claim depends upon the circumstances. At *PI* §164, he notes, regarding the possibility that someone who is casting her eyes over a text may not in fact be reading, that "in different circumstances, we apply different criteria for a person's reading". And as Phillips (1978) notes, evidence is only criterial evidence in certain circumstances. This circumstance-dependence is taken by Baker to imply that C-relations are defeasible: the

claim to knowledge based upon a particular criterion may fail in just those cases where in fact the evidence is *not* a criterion for the claim. That is, if the criteria for a claim are satisfied (for example, someone may be manifesting pain-behaviour), then we may not be able to make the claim that she is in pain, simply because we learn that in fact she was acting at the time: we can have true evidence for a false claim.

However, although C-relations are weaker than entailments (though they can still support certainty), they are not to be confused with inductive relations. In any case in which criteria support a claim, the claim is made certain; in contrast, were it an inductive relation, the claim would only be made probable to the degree that its underlying inductive generalisation were supportable. Such an inductive relation is necessarily based upon experience, and so cannot conform to the requirement that C-relations be knowable *a priori*. Indeed, in order to keep clear the distinction between C-relations and inductive relations, Wittgenstein introduced the notion of *Symptoms*, or *S-relations* to describe the latter. S-relations may form part of the sense of a term, and indeed may be evidence to support the use of a term for example in making a knowledge-claim about the world; but it does not form part of the evidence that makes the claim certain.

Another aspect of the C-relation is that it is an *internal* relation: that is, the meaning-relation itself, though dependent upon circumstances, is not to be decomposed into a pair or relations with some independent third entity. So if x is C-related to y, then x is taken directly as evidence for y, without first checking on the appropriateness of the circumstances or the possibility of defeaters. That is, the "transition" from the criteria to the facts of the matter is *a priori*.

Now, the possibility that meaning relations may be defeated in the manner of C-relations also seems inherent in the formulation of conditional constraints. In their base form, conditional constraints are as follows.

$$p \Rightarrow q \text{ given } r$$

That is, the relation between p and q does not hold universally. In particular, it does

not hold in situations where r does not hold. Let us call such situations $\neg r$-situations. Constraints such as these, however, do hold in what we may call r-situations. Now, this aspect of conditional constraints gives rise to a kind of defeasibility.

First, let us note the character of a conditional constraint in r-situations. Within situations of this type, the conditional constraint behaves just as an entailment. That is, within this type of situation, if p is true then q must be true.[1] So, for example, if an agent never encountered a $\neg r$-situation, the agent's behaviour would be such that one could not discriminate the nature of the relation: one could not distinguish a conditional constraint from something akin to an entailment.

This likening of conditional constraints to local versions of entailments renders them different from certain other defeasible relations and, in particular, probabilistic relations. A probabilistic relation holding between p and q would serve to indicate that in any given situation there is a certain probability associated with the outcome q given the outcome p. That is, if the relation between p and q is truly a probabilistic one, then there can be no *a priori* grounds for discriminating between situations in which both p and q hold and ones in which just p holds. That is the nature of genuinely probabilistic relations. Conditional constraints, however, are quite different. Barwise (1985) considers the relation between his daughter Claire rubbing her eyes and her being sleepy. The question that an analysis begs is whether this relation is unconditional, conditional or probabilistic.

If the relation is unconditional then in all circumstances in which the antecedent condition is satisfied (Claire rubbing her eyes) the consequent condition (Claire being sleepy) will also be satisfied. No situation can prove to be exceptional in this regard. The fact that there are other situations, ones in which there is an abundance of pollen in the air, in which Claire rubs her eyes and yet is not sleepy is, then, a demonstration that the relation between eye-rubbing and being sleepy is not of this unconditional nature.

[1] It should be noted that our discussion is more properly centred on what Barwise & Perry (1983) call *necessary* conditional constraints. Whenever we talk of conditional constraints, it is this kind of constraint we mean.

Conditional Constraints	Barwise & Perry, Sense Generation
Unconditional Constraints	Kripke, Putnam

Table 2.1: Choice 1: Meaning Relations as Conditional or Unconditional Constraints.

One possibility, then, is that the relation is a conditional one. If this is the case, then there will be certain types of situation in which the relation holds between eye-rubbing and being sleepy and certain types of situation in which it does not. Notice that this is different from the picture we would obtain if the relation was truly probabilistic. For then it would be the case that the relation would hold in certain situations and not in certain others. If the relation is conditional then it holds in certain *types* of situation and not in certain other *types*. That is, there is the possibility of classifying those situations according to their type in a way that is not possible with probabilistic relations. That is, with a conditional constraint we have the possibility, in principle, of examining the current situation, of determining its type, and consequently determining whether the relation holds or not. With a truly probabilistic relation, we have no such possibility. We may only examine actual outcomes. This difference between conditional relations and probabilistic ones, a difference relating to the former's *a priori* nature and the latter's *a posteriori* one, is an important one.

The choice, then, in analysing meaning relations in terms of constraints, seems clear and some positions are summarised in Table 2.1. Though discussion of Sense Generation will come in Chapter 3, the choice it adopts is also indicated.

1. **Meaning Relations as Conditional Constraints:** allows meaning relations to give rise to apparent defeasibility by the fact that they hold in certain types of situation and not in certain other types.

2. **Meaning Relations as Unconditional Constraints:** disallows meaning relations from exhibiting defeasibility.

2.1.2 One or More: the Unitary Hypothesis

The choice we describe in this section is between viewing the informational content of the uses of a single unambiguous word as arising from *one* meaning relation or *more* than one meaning relation. That is, whether we see the relation between the uses of a word and the information they convey as effected by a single constraint or by many such constraints. In many ways, the question reduces to that of deciding whether the uses of a word betray a single set of conditions which underly its use, or whether it betrays many such sets of conditions. In perhaps a trivial sense, most would agree that several meaning relations underly *all* the uses of a given word. Cases of metonymy and metaphor, it is standardly assumed, are cases where the word has a different meaning from the conventional. However, our concern is not so much with just these "peripheral" uses. Rather our question is whether the conventional uses of single words belie single or multiple constraints. In the case, for example, of those uses of *lemon* which refer to real lemons, the issue is whether such uses belie one informational constraint or many. As we shall see, which choice we take here depends on our previous choice.

One view which might take meaning relations as unconditional constraints is the one which we alluded to above, that of Kripke and Putnam. The major concern of their articles *Naming and Necessity* and *Is Semantics Possible?* is precisely with these conventional uses of words. Now, on their view, it is assumed that the meaning of a word must be expressible in terms of necessary and sufficient conditions on word use. Such conditions must, therefore, be ones so as to capture all the possible facts of word use in all situations of use. Ignoring the difficulties associated with this view, it is clear that such a conception allows no motivation for the view that the conventional uses of a word belie more than one meaning relation: to what possible use could any additional meaning relations be put? If each individual meaning relation specifies necessary and sufficient conditions, then the theorist has no need of any further meaning relations.

So, if we postulate meaning relations which are unconditional, which hold in all situations, then there can be no obvious motivation for specifying more than one such meaning relation to explicate the conventional uses of a given word. However, to be

thorough, we will note that to postulate multiple unconditional constraints to explicate the uses of words is, technically, an option.

Now the view which treats conventional uses as arising from conditional constraints does have a more meaningful option here. Firstly, one could assume that only one conditional constraint belies the conventional uses of a word. However, in this case, there may be problems in dealing with counterfactual conditionals. Let us suppose that the particular constraint is as follows.

$$s \models \ll \Rightarrow, S_1, S_2; + \gg$$

So, S_1 might be the type of situation in which an entity had the property of having atomic number 79 and S_2 might be the type of situation in which the word *gold* applies to that entity. Now, an important fact about the constraint above is that it is a component of situation s and of all smaller situations, but not of ones which are not included in s. That is, there are situations in which the constraint does not hold. In order to explain the facts of language use, the theory would have to consider the possibility of cases where the current situation was not of type S_1 and yet where the word *gold* does apply to entities which are components of that situation. It would not be enough to simply assert that the current situation was not included in s, for this fails to offer an explanation of the facts of language use in precisely situations of this sort. What we require is a theory that explains the facts about language use in *all* situations in which a word may be used. So, the only meaningful option would be to alter our definition of S_1 so that the constraint correctly predicted the facts of language use in *all* situations in which *gold* could be used and to relativise the constraint to a larger situation s'. But, since such situations seem boundless in number and nature, this option would appear to reduce to the first, one in which meaning relations are taken as unconditional constraints.

However, another possibility is to view the conventional uses of a given word as involving more than one meaning relation, more than one conditional constraint. By these means we can avoid the pitfalls associated with viewing meaning in terms of single conditional

	Conditional Constraints	Unconditional Constraints
One	??	Kripke, Putnam
More	Sense Generation	??

Table 2.2: Choice 2: Single or Multiple Meaning Relations. Constraints.

constraints, the pitfalls associated with counterfactual conditionals. Now, the argument is that p involves q given r and that in certain circumstances, those in which r does not hold, other meaning relations hold: p' involve q given r'. Clearly, a challenge for such an approach is to spell out the relation between p and p' and between r and r'. But this possibility of considering multiple conditional constraints as underlying the conventional uses of words is, nonetheless, one which circumvents arguments from counterfactual conditionals.

We now have a four-way choice: between unconditional or conditional constraints; and between one or many of these. This is summarised in Table 2.2.

2.1.3 Are Meaning Relations Descriptional?

In answering this question, we will first consider one notion of meaning relations, that deriving from the Fregean notion of sense. At a later stage (section 5.1.1 and section 5.3) we will have a little more to say about Frege's view of senses but, for the purposes of this section, we will concentrate more on how this notion has been interpreted and used in the psychological literature on word meaning. Finally, we will relate this to the notion of informational content in Situation Theory which we discussed in Chapter 1.

The standard Fregean picture of sense prevalent in psychology is as follows: words have two aspects to their meaning, sense and reference. A proper name like Aristotle, for example, may have the sense "the teacher of Alexander" and has, of course, as referent, Aristotle himself. Sense is taken to determine reference. That is, it is the satisfaction or otherwise of descriptions such as "the teacher of Alexander" that determine precisely which individual is the referent of the name. Senses of words contribute to the senses of

sentences (thoughts) by the principle of compositionality. Further, senses are parts of the common language and are grasped by every competent user of the language. Finally, though Frege himself never stated that senses are descriptions, it seems a commonplace view in psychology.

Johnson-Laird (1987), for example, adopts this view of senses and, equating them with the meanings of words, he focusses on their mental representation. In particular, he claims that "there are entries in the mental lexicon that allow ready access to the information that an individual has about the sense of a word" (Johnson-Laird, 1987; p. 205). As to the question of whether such meaning relations can be described, assuming that this does not follow from the fact that they can be represented, the answer from Johnson-Laird seems to be equivocal. While accepting that there may be some words, natural kind terms, for example, whose senses are not amenable to description and, consequently, receive only a partial entry in the lexicon, he points to the fact that for other words (his examples are words such as *plate* and *vase*) their senses can be described. Indeed, these descriptions can form part of the entries in the mental lexicon. So, Johnson-Laird's answer is that, meaning relations may be descriptional in principle but, in practice, it depends on the particular word.

Another area of psychology in which we come across the notion of sense is Clark's work on Sense Creation (Clark, 1983). Clark's concern is with the senses of novel uses such as that of *porch* in *the paper boy porched the newspaper.* So Clark's concern is only indirectly related to the issue of whether or not senses can be described. But Clark concurs with Johnson-Laird. Senses can, in principle at least, be described. Indeed, Clark's ability to explain the sense of denominal verbs such as *to teapot* depends on this fact.

In many respects, the notion of sense employed in the psychological literature is akin to the notion of informational content. Barwise (1989) draws a distinction between three uses of *meaning*: when it is used to mean *sentence meaning*, when it used to mean *content* and when it is used to mean *user meaning.* The notion of content here refers to the meaning attached to "the event of A's using it [a sentence] to say something to

B, a specific action which means something" (Barwise, 1989; p. 62). That is, content refers to the meaning of an act of uttering some sentence, namely, at least at one level, the conveying of information. And this seems to be precisely what the psychological notion of sense is getting at, not the notion of sentence meaning to which senses are often tied in the philosophical literature. So, the meaning of *A* saying to *B* "The paper boy porched the paper" is equivalent, in some sense, to the situation-theoretic notion of the informational content which that utterance conveys.

So, at least with respect to the conventional uses of natural kind terms, the psychological literature seems consonant with the idea that these senses are not descriptional. The philosophical literature, too, seems to espouse the same view. Kripke, for example, in his *Naming and Necessity* rejects the Fregean picture of sense for natural kind terms and claims that they are more like proper names in having solely a referent or denotation.

However, the reasons for such a rejection are largely due to the fact that such words are deemed to have a single meaning. That is, there is a single meaning relation or sense underlying all their uses. And it seems a fact of natural kind terms that all their uses cannot be given one single description. Hence, the view that arises is that natural kind terms either do not have sense or that such senses are non-descriptional. So, one option, that of taking meaning relations and senses to be non-descriptional in certain cases, appears to be a corollary of the previous choices.

By allowing a word to have more than one meaning underlying its conventional uses, this problem is circumvented and one can claim that such meaning relations, even in the case of natural kind terms, are descriptional. The results of this section are tabulated in Table 2.3. Though we haven't discussed Sense Generation, we will indicate where it stands on the issues raised in order to act as something of an appetiser for the next Chapter.

	Conditional Constraints	Unconditional Constraints
One	??	Kripke, Putnam (Non-descriptional)
More	Sense Generation (Descriptional)	??

Table 2.3: Choice 3: Are Meaning Relations Descriptional?

2.2 Lexons

LEXON is the term we reserve for that component of entries in a mental lexicon which represents so-called "semantic" information. That is, information which specifies either totally or partially the meaning of a word on given occasions of use. The question of whether a particular LEXON adequately captures the facts pertaining to all the uses of its corresponding word or merely some subset thereof, is one we discuss in section 2.2.2. Before we discuss the choices which we should consider for LEXONs, let us turn to a discussion of the nature of a mental lexicon.

A mental lexicon is deemed to be the locus of those mental representations which express the linguistic properties of words. So, for instance, each word will be deemed to have a mental representation which encodes facts concerning the word's syntactic category, its subcategorization, its phonology, morphology and its semantics. It is the nature of the semantic component of entries in a mental lexicon with which this thesis is primarily concerned.

Clark (1989) offers the following summary of theories such as these.

> Every word has a lexical entry in memory that pairs a phonological shape, like */dog/*, with a conventional meaning, like "canine animal". The conventional meaning is really a brief, partial description of some aspect of the world. All the words taken together form a list called the mental lexicon. When we need a word, we search this list for a word with the right conventional meaning. And when we hear the phonological sequence */dog/*, we search the list for that shape and retrieve its conventional meaning. Although this may be somewhat of a caricature of the dictionary theories, it

isn't far wrong.

(Clark, 1989; p. 2)

Now Clark, in a series of papers (Clark, 1983, 1989; Clark & Gerrig, 1983; Morrow & Clark, 1988), has presented a number of reasons why such dictionary theories are inadequate as theories of the mental lexicon. Indeed, we will discuss his own suggestion later (Chapter 3). For our current purposes, however, we simply need note that these theories and Clark's alternative both presuppose the existence of a mental lexicon. Clark's concern is to do with how various "lexical and conceptual possibilities" are to be represented: for example, whether the possibilities for the hues which *red* may denote are to be represented in terms of a static "dictionary" or whether they are the products of some generative lexical process. But the lexical process he suggests, itself takes as input entries in a mental lexicon.

Johnson-Laird (1987) is perhaps more explicit. His conclusion regarding the mental representation of the meanings of words is that information about word senses is located in "entries in a mental lexicon" (Johnson-Laird, 1987; p. 205). That is, a mental dictionary that contains entries in which the senses of words are represented, and these senses may be more or less complete. For example, if the language user is ignorant of the full sense of the word or if the word is a theoretical term, then the represented senses may be incomplete. Johnson-Laird suggests that the representation of the senses of natural kind terms may be relatively incomplete in this way. At any rate, what is clear is that within the psychological literature on word meaning, it is a commonplace assumption that information pertaining to word senses is represented in a mental lexicon. We will not challenge this assumption in any great detail, though our concern is to present both a rejection of dictionary theories and an alternative. This we do in Chapters 3 and 4. Later in this chapter, however, we will consider in more detail the representational assumptions we make in positing entries in a mental lexicon.

2.2.1 One or More

The choice we have here is precisely how many LEXONs we postulate in order to represent the variety of word senses. Take, for example, *light*. This has at least two word senses: "light-weight" and "light-coloured". So it is that we can say of a raven's feather both that it is, and that it is not, light. Such a condition Quine (1960) takes as an indicator of the ambiguity of *light*. Now it is standardly assumed that, in the case of ambiguous words, their different word senses are represented in different entries. That is, we would expect two LEXONs for the single word *light*. Indeed, such a method of solution characterises the dictionary theories of word meaning in which all the different senses a word may have are simply listed, each sense warranting a different entry. Such a solution, however, cannot be general.

Zwicky & Sadock (1975), in their discussion of ambiguity tests, contrast cases of ambiguity with cases of what has variously been termed vagueness, indeterminacy or generality, this last being Quine's term. The choice is, as Zwicky & Sadock point out, between postulating distinct underlying semantic representations in the case of ambiguity and postulating single representations that correspond to several distinct states of affairs in the case of generality. Take the case of the word *cousin*. This word can be used to refer both to females and males, but the word does not appear to be ambiguous between these possibilities. It is not that the word has two underlying representations, one corresponding to male and one to female cousins. Rather, it seems that the word does not specify the sex of the referent(s). That is, it is general with respect to this possibility. Similarly, in the case of *dog*. The more frequent reading of *dog* holds that it is general with respect to whether the referent is male or female. However, there is also a reading "male canine" which is clearly not general.

The issue, then, for any theory of word meaning is how it treats the variety of senses words appear to have. Take the case of the Lion Puzzle of Chapter 1. Here there appear to be two senses of *lion*. One which appears to apply to statues of lions, one which does not. So is *lion*, for example, to be accorded two LEXONs or merely one? Do we treat such uses as manifestations of an underlying ambiguity? Or are they manifestations of

Number of LEXONS	
One	More
Sense Generation	Dictionary Theories, Sense Selection

Table 2.4: Choice 5: The Number of LEXONs for Non-ambiguous Words

an underlying generality?

We will consider these options in detail in Chapter 3 but it is worth noting that there is a further option concerning the number of LEXONs we postulate. Suppose that there are good reasons for believing *lion* not to be ambiguous. Such a conclusion would lead us to postulate one LEXON for *lion*. Now, by assuming that *lion* is general with respect to the possibilities suggested by its uses, we would conclude that the LEXON for *lion* is unspecified with respect to many properties: whether animate or not, whether made of flesh and bone or not, whether golden brown or stone grey, etc. Indeed, such a LEXON would be considerably inspecific. An alternative, however, which we discuss in Chapter 3 and which we raise in the next section, is to view the LEXON as embodying default information and the various senses which the word may have as arising from some further process, one which is clearly non-monotonic.

The choice for the number of lexons then concerns whether we treat the various senses of a case like *lion* as manifesting ambiguity or not, as manifesting one or more LEXONs. This simple choice is summarised in Table 2.4 and, again, we indicate where Sense Generation stands. It is worth re-iterating, though, that the table does not indicate a stand with respect to all words, only with respect to those words which seem not to be ambiguous. Thus, one option is to treat all cases where words have more than one sense as instances of ambiguity. Sense Generation, while accepting that many words are genuinely ambiguous, holds to the view that many non-ambiguous words have multiple senses. It is these words that should receive only one LEXON.

2.2.2 The Content of Lexons

The descriptive content we associate with lexons will crucially depend on the number of lexons we postulate. Let us, by way of example, consider the case of *mother* from section 1.2.3. As Lakoff (1987) points out there appear to be distinct senses of *mother* associated with the cognitive models he proposes underlies its use. These are the birth, genetic, nurturance, marital and genealogical models. Now, let us consider each of the above choices in turn: ambiguity, generality and default.

Ambiguity: This choice would entail us treating *mother* as an ambiguous word. For each of the five different senses of *mother* there would be a distinct LEXON. The contents of these lexons would be treated as if they were as independent from one another as the different senses of *bank* or *light*. The content of each LEXON would be that content which would pick out each type of mother: the LEXON corresponding to the birth model would specify content that would refer to adult females who have given birth to a child; the LEXON corresponding to the nurturance model would specify content that would refer to adult females who have nurtured a child; and so on. That is, the content of a LEXON is determined by the different senses we assume a word to have.

In the case of the Lion Puzzle, this choice would entail there being at least two LEXONs for *lion*. The content of one would be such that it would apply solely to real lions; the content of the other would be such that it would apply solely to statues of lions.

Generality: This choice would entail that we treat the different senses of a word as different manifestations of the same general or inspecific LEXON. Consequently, the content of the LEXON must be general with respect to all possible senses. In the case of *mother*, the LEXON would not specify a genetic relationship, a birth related relationship, a nurturance relationship, a marital relationship or a genealogical relationship. In informational terms, the LEXON would amount to the disjunction of all these relationships, provided these particular relationships exhaust the different senses of *mother*.

In the case of the Lion Puzzle, this choice would entail there being one single LEXON

	General/Inspecific Content	Default Content
One LEXON	Generality Option	Sense Generation
More than one LEXON	Weak Sense Selection	Strong Sense Selection

Table 2.5: Choice 5: The Number and Content of LEXONs

for *lion*. Its content would be inspecific with respect to distinguishing properties of real lions and statues of lions. That is, the LEXON would not specify the property of animacy since statues are inanimate; it would not specify "made of flesh and blood" since statues are not; and so on.

Default: This choice would entail that we treat the different senses of a word as being derived non-monotonically from a LEXON which specifies the default content associated with senses of the word. In the case of *mother*, for example, the LEXON would specify the default sense of *mother*, presumably one that satisfies all the models which Lakoff proposes. In the case of other senses of *mother*, their content would be derived non-monotonically from the content of this LEXON.

In the case of the Lion Puzzle, this choice would entail there being one single LEXON for *lion*. Its content would express the default sense of *lion*, presumably "real, animate lions". The content of senses such as "statue of lion" would then be derived from the content of this LEXON.

Later (section 3.1), we will discuss reasons why the ambiguity and generality options are not appropriate for the kinds of puzzles considered in Chapter 1 and why the last option, the default option, is preferable. For now, however, it is enough to note the various choices and their interdependence. Again, we indicate where the theories we will discuss later, Sense Generation and the Weak and Strong forms of Sense Selection, stand with respect to this choice.

2.2.3 The Relation Between Senses and Lexons

Even given the earlier choices there is still freedom in defining the relation between senses and LEXONs, especially in terms of their content and number. In particular, those theories which postulate more than one LEXON in order to capture the observed diversity of the senses of words, are relatively unconstrained in determining precisely what number of LEXONs they postulate. Similarly, and concomitantly, they are unconstrained in the contents they postulate for such LEXONs. Indeed, this lack of constraint may lead to the claim that such theories require furnishing with explanatory principles to determine precisely these issues of content and number. We will not dwell on this matter. However, one aspect of the relation between senses and LEXONs upon which we have already touched is that of the relations between their contents and, in particular, whether the relation is monotonic or otherwise.

The Generality option of the previous section is one that is forced by standard views of word meaning. According to the discussion of Zwicky & Sadock (1975) we are led to believe that there are only two possibilities for word senses: either the word is ambiguous, having multiple senses; or it is general having an inspecific sense. Now, postulating an inspecific LEXON entails that the addition of particular sorts of information is strictly monotonic. Take *cousin* for example. As we have already noted, this appears to be general with respect to sex. Consequently, any further specification of the sense for *cousin* with the information male or female is monotonic. The added information contradicts nothing. Indeed it is to avoid such contradiction that the claim of generality is made.

Suppose that we assume *bird* to have the sense "flying animal". This plausible suggestion is confounded by the fact that there are certain birds which do not fly. That is, the facts pertaining to these birds contradict the assertion contained in the sense. Consequently, to avoid such a contradiction, the claim of generality is made: since there are birds which fly and birds that do not, *bird* must be general with respect to these alternatives. *Bird* cannot mean a flying animal.

Notice that the argument proceeds from the possibility of contradiction to a revision of

	Monotonic	Non-Monotonic
One LEXON	Generality Option	Sense Generation
More than one LEXON	Sense Selection	??

Table 2.6: Choice 6: The Relation Between the Content of Senses and LEXONs

the sense and/or LEXON for *bird*. However, there is nothing to suppose that contradictions of this sort do not or should not occur. Indeed, in the case of the Lion Puzzle, the possibility of contradiction seems unquestionable. Here, it seems, it is perfectly possible to say of lions that they are animate creatures and that an inanimate statue of a lion is nonetheless, in some sense, a lion. So, it seems, that in some cases at least, entities to which a word applies contradict the assertions contained in the sense of that word.

In general, then, there seems to be no *a priori* reason why one would wish the relation between the senses a word may have and its LEXON to be a monotonic one. It merely seems a commonplace assumption. The possibility that the relation could be non-monotonic, however, is one which rejects the Generality option, at least as a viable move for all cases. Indeed, if one accepts that the content of a sense may be non-monotonically derived from the content of a LEXON then one may allow LEXONs to encode information pertaining to the default sense, and yet still allow for any observed diversity of senses. The possible relations between the contents of senses and LEXONs is detailed in Table 2.6.

2.2.4 Describing Senses and Lexons

In this section we shall briefly consider the relationship between the kind of language in which we will represent senses and LEXONs, and claims concerning mental representation.

In assuming that senses and LEXONs may, in principle, be descriptional, we are committed not only to their description but to a particular manner of description. Consistent with much work both in the linguistic and psychological literatures, we will describe

senses and LEXONs in terms of what are variously called feature-structures, directed acyclic graphs, attribute-value pairs, etc. Much of the detail of these we will leave at an intuitive level. However, in the following, we explore one formal language in which one may express facts concerning feature structures. A language in which we may describe the content of senses and of LEXONs provides, among others, a means by which we may classify theories of senses and LEXONs. Indeed, in Chapter 3, we will classify a number of different such theories. The formalisation that we offer is similar, though not identical, to Johnson (1988).

Johnson (1988) introduces and later formalises, in terms of Attribute-Value Logic (AVL), a language for describing feature structures. There are several reasons for choosing such a language. One concerns its expressive power and since, as Shieber points out, this is likely to equal that of a Turing machine, we should be able to compute anything that we both want to compute and that is computable. Another reason is the widespread use of such languages in linguistic theories, particularly the mono-stratal, lexical theories such as HPSG (Pollard & Sag, 1987), CUG (Uszkoreit, 1986), etc. It is not inconceivable that an implementation in terms of such a language might one day be reconciled with one or other of these approaches.

There are several components to the basic language which defines Attribute-Value Structures or (*AVS*s).[2] An (*AVS*) is a quadruple as follows: $AVS = \langle \text{ATTR}, \text{VAL}, \text{LAB}, \delta \rangle$ where ATTR is a set of atoms called attributes, VAL is a set of atoms called values and LAB is a set of atoms called labels. Further, ATTR, VAL and LAB are disjoint. δ is the partial function δ: $\text{LAB} \times \text{ATTR} \mapsto \text{LAB} \cup \text{VAL}$.

Consider the following *AVS* as an example.

Let $\text{ATTR} = \{a_0, a_1, a_2\}$, $\text{VAL} = \{v_0, v_1, v_2\}$, $\text{LAB} = \{l_0, l_1, l_2\}$ and δ be as follows.

[2]This characterisation of an *AVS* differs from that of Johnson (1988), where an *AVS* is defined as a triple $\langle F, C, \delta \rangle$, where $F = \text{ATTR} \cup \text{LAB} \cup \text{VAL}$; $C = \text{ATTR} \cup \text{VAL}$, and δ is the partial function δ: $F \times F \mapsto F$, subject to the restriction that $\delta(c, f)$ is undefined for all $c \in C$ and $f \in F$. Johnson's characterisation allows what we call attributes to act as what we call values and vice versa; our characterisation explicitly disallows this. However, the difference is minor and does not affect the points we make later.

$$\delta(l_0, a_0) = l_1 \quad \delta(l_1, a_0) = v_0 \quad \delta(l_2, a_1) = v_0$$
$$\delta(l_0, a_1) = v_1 \quad \delta(l_1, a_2) = l_2 \quad \delta(l_2, a_2) = v_2$$

Such an attribute-value structure can also be represented diagrammatically in the usual manner (cf. Shieber et. al., 1986) as follows.

$$\left[\begin{array}{ll} A_0 : & \left[\begin{array}{ll} A_0 : & v_0 \\ A_2 : & \left[\begin{array}{ll} A_1 : & v_0 \\ A_2 : & v_2 \end{array}\right] \end{array}\right] \\ A_1 : & v_1 \end{array}\right]$$

We assume the usual subsumption ordering on such structures. Identity of attribute-value structures holds if and only if identity holds of the corresponding partial functions. Two attribute-value structures AVS and AVS' are ordered by subsumption ($AVS \sqsubseteq AVS'$) if and only if the corresponding partial function δ' is a *extension* of δ.

In the discussion of the later chapters, we will be concerned to describe the senses and LEXONS associated with given words. Accordingly, we need a convenient way of notating the various lexons for a given word and senses for a given LEXON and we will notate these by indices. For example, we will notate the AVS associated with the jth LEXON of the ith word as $AVS(L_{ij})$, and the kth sense associated with that LEXON, as $AVS(S_{ijk})$. Consequently, for each unambiguous word j will assume just one value; for ambiguous words it will assume more than one value.

So far in this section we have been concerned to illustrate the kind of language with which we may describe senses, LEXONS and the relations between their contents. The classification we offer in Chapter 3, will be based on this kind of language, though, it will be clear, that it does not depend on any particular version. However, the main purpose of this section was to reflect on the claims that one might be tempted to make on the basis of phrasing a psychological theory in the terms of a language such as the one described. It is to a consideration of some of these claims that we now turn.

The content that we postulate for senses is primarily intended to serve a classificatory role. Senses are seen as classifiers of linguistic behaviour; of when we may or may not use a given word to refer to a given entity. To describe senses in terms of attributes and values is, then, to effect a classification of linguistic behaviour in such terms. So, for instance, if we claim that one sense of *lion* is classified by the attribute-value pair

ANIMATE: +

then the claim would be that such a sense may apply to entities which are indeed animate. Entities, that is, possessing the property of animacy. So senses classify linguistic behavior in terms of the properties that entities may possess. Unfortunately, when we consider LEXONs, their classificatory role is more complicated.

Even though we may describe LEXONs in the same language with which we describe senses, the claims we attach to LEXONs must necessarily be more carefully made. LEXONs play a dual role in a psychological theory. Firstly, they play a role similar to that proposed for senses: that is, LEXONs are deemed to classify behaviour. Secondly, they are deemed to correspond to some mental representation. Let us consider each aspect in turn.

The assumption of a mental lexicon ties one to the view that the content of entries in that mental lexicon has some privileged role compared to other content. One may compare such a view of the content of LEXONs with syntactic phenomena. A word's morphology, for instance, may be represented in the lexicon in infinitival form. So, under the entry for the word *take*, the morphology will be represented as *take*. Other morphological forms will then be generated from this base infinitival form by means of productive morphological rules. So the one base form assumes a privileged status compared to the other derived forms and this difference in status, if interpreted as a psychological claim, will manifest itself in terms of differences in behaviour. And so it is with the semantical entries of a mental lexicon or LEXONs. If we assume that certain content is represented as LEXONs and certain other content is derived from this, then we may expect there to be differences in behaviour that accords with this difference in

content. Similarly, if we observe differences in behaviour that accords with differences in semantic content then we may expect these different contents to be expressed differently in our theory. Some content may be expressed as the content of a LEXON, other content may be seen as derived content. Now, should we chose to describe these contents in terms of attributes and values, then we have said no more or less than such and such behaviour may be classified in terms of these attributes and values. A problem arises, however, when we claim that these LEXONs are mentally represented.

The claim that LEXONs are mentally represented amounts to the interposition of a third entity between two we have already discussed: behaviour and attributes and values, or properties. And in this interposition lies great scope for what appear to be some erroneous claims concerning mental representation.

One such erroneous claim would be the claim that the language in which we, as theorists, represent the content of LEXONs is one and the same language as that in which this content is mentally represented. For the facts concerning mental representation do not appear to be consistent with what we would expect if these mental representations are as our formal language is. Symbolic systems, whose operations are defined in terms of a formal language, perhaps like that we have earlier described, typically do not exhibit the properties we have come to associate with mental representations. Facts such as the graceful degradation of human memory, for example, seem incompatible with a view of memory based on standard symbolic computation. Indeed, in many respects, the computational properties of a distributed representation account of word meaning would appear to more closely resemble those of human architectures. So, at any rate, we do not claim that the attribute-value structures by which we describe LEXONs are represented in these very terms in the head. Rather, our claim is that these attribute-value structures classify mental representations.

It is a standard result of computational theory that any computable function may be computed in an arbitrary number of ways. So, for example, we may compute the function which maps 2 and 3 to 5 in an indefinite number of ways. Indeed, it is surely no different in the case of the theory of word meaning we will outline later. It, we

will claim, suggests that the function between words and their uses may be computed in a certain way. It does not, indeed it cannot, suggest that this is the only way of computing this said function. And so the claim cannot be that this is the way in which such a function is mentally computed. Rather, the claim is weaker: that the function which we describe, between words and their uses, is mentally computed and that the content of LEXONs which we postulate and which is implicated in this function classifies mental representations. So, we assume that the content of LEXONs provides a means for individuating mental states. Someone with a different content for the same LEXON would be in a different mental state. Of course, the means by which we, as theorists, may individuate mental states is not something we wish to illuminate. Indeed, such a task seems Herculean. All we wish to establish is the precise claims that one may attach to our postulation of particular LEXONs with particular contents. And the claims are twofold. Firstly, that this particular content can be taken to classify behaviour. Secondly, that this particular content classifies mental states. We say nothing more.

It is clear, then, that commitment to this kind of formal descriptive language in no way constitutes a commitment to decompositional theories of word meaning though neither does it constitute a rejection of such theories. The issue of whether word meanings may be mentally represented in terms of decompositions or meaning postulates is, then, quite independent from the issue of whether such representations can be classified in terms of LEXONS. From the perspective adopted in this thesis, decompositional and meaning postulate theories are indistinguishable since both can achieve the same formal classifications of supposed mental states. They are only distinct when they are interpreted as claims about the mental states themselves.

We finish this section by offering some justification for the kind of language we have chosen. Indeed, our choice may seem ill-judged especially given our comments of section 7.2.1 that, in the light of arguments concerning coherence, such a language is particularly ill-suited for a theory of concepts. However, our choice is motivated primarily by the need to compare the approach we develop later (Chapter 4) with prototype theory. Most of the explicit formulations of prototype theory are phrased in terms of an attribute-value language. So, in order to effect the comparison with prototype theory

in Chapter 7, we will also adopt a simple attribute-value language.

2.3 Conclusion

In this chapter we have examined a number of different choices that a theory of word meaning may take. The first concerns whether we view word meanings in terms of conditional or unconditional constraints. The second concerns whether we assume that the various uses of a word betray one or more constraints. Thirdly, we focussed on the issue of whether word meanings are descriptional. We have suggested that this choice is determined in part by the previous two. Viewing the uses of a word as betraying a single unconditional constraint would support Kripke's conclusion that the meanings of certain words, natural kind words for example, are non-descriptional. If we regard the uses of a word as betraying many conditional constraints, however, it seems that one may still hold to the view that word meanings are descriptional even in the case of natural kind words. Our fourth and fifth choices reflect the number and content of entries in a mental lexicon that we postulate for various words. The choices determine whether we choose to treat certain words as ambiguous or general. An alternative is to view LEXONs as embodying default information. This then reflects on the final choice which is the relation between the content of senses and LEXONs. We mentioned the possibility that one can see this relation of content as being either monotonic or non-monotonic.

Our final concern in this chapter was to put some distance between the formal language in which we describe the contents of senses and LEXONs and a claim about mental representation. Our position is that while senses must be seen as classifying linguistic behaviour, LEXONs must be seen to perform a dual classificatory role. They classify both behaviour and mental states. The precise manner in which this classification is effected is not a topic we address, though an account of this must form a part of any complete cognitive scientific theory of word meaning. With this much in place we are now in a position to consider a particularly influential argument, found in the psychological literature, which addresses the nature of senses and argues for a form of Sense Creation.

It is this which forms the starting point of the next chapter and our exposition of Sense Generation.

Chapter 3

Sense Generation

> One Sunday night I happened to walk for some fifteen paces next to a group of six drunken young workmen, and I suddenly realised that all thoughts, feelings, and even a whole chain of reasoning could be expressed by that one noun, which is moreover extremely short. One young fellow said it harshly and forcefully, to express his utter contempt for whatever it was they had all been talking about. Another answered with the same noun but in a quite different tone and sense — doubting that the negative attitude of the first one was warranted. A third suddenly became incensed with the first and roughly intruded on the conversation, excitedly shouting the same noun, this time as a curse and obscenity. Here the second fellow interfered again, angry at the third, the agressor, and restraining him, in the sense of "Now why do you have to butt in, we were discussing things quietly and here you come and start swearing." And he told this whole thought in one word, the same venerable word, except that he also raised his hand and put it on the third fellow's shoulder. All at once a fourth, the youngest of the group, who had kept silent till then, probably having suddenly found a solution to the original difficulty which had started the argument, raised his hand in a transport of joy and shouted ... Eureka, do you think? Found it? Found it? No, not Eureka at all; nor did he find anything; he repeated the same unprintable noun, one word, merely one word, but with ecstasy, in a shriek of delight — which was apparently too strong, because the sixth and the oldest, a glum-looking fellow, did not like it and cut the infantile joy of the other one short, addressing him in a sullen, exhortative bass and repeating ... yes, still the same noun, forbidden in the presence of ladies but which this time clearly meant "What are you yelling yourself hoarse for?" So, without uttering a single other word they repeated that one beloved word six times in a row, one after another, and understood one another completely.
>
> *(Dostoevsky, The Diary of a Writer)*

Having established some choice points for theories of meaning and sense, in this chapter

we characterise one option called *Sense Generation.* Sense Generation admits of a number of distinct approaches though they are all similar in spirit, if not in detail, to the view that Herb Clark espoused in his 1983 paper, *Making Sense of Nonce Sense*, and that Clark & Gerrig (1983) called *Sense Creation.* Sense Generation is indeed motivated in part by Clark's considerations concerning nonce sense and the senses of contextual expressions. However, what distinguishes Sense generation from these expositions of Sense Creation is the treatment of conventional uses of words, the traditional domain of theories of meaning and sense.

We begin by examining Clark's arguments in favour of Sense Creation. Clark presents these by way of a rejection of what he calls the *sense selection* assumption. This is a standard assumption of parsing which he claims to be invalid for the parsing of contextual expressions. We differ from Clark in that our concern is not simply with contextual expressions but also the conventional uses of words. In our analysis, we will consider two distinct types of Sense Selection, Strong and Weak, and possible variants of these. Whereas Clark's original objections apply to Strong Sense Selection, it seems that Weak Sense Selection may obviate these. However, arguments against both forms of Sense Selection can be adduced from the psychological literature and we will develop these in some detail. Most importantly, these arguments strongly suggest a number of attributes we would expect a theory of sense to have. Such a theory, we will argue, is Sense Generation.

We then turn to a brief outline of some of the intuitions behind Sense Generation and of its key claims and these will be illustrated by an example. One aspect we will not consider straight away is the mechanism by which senses are generated. Rather, our attention will turn first to some of the objects which play a role in the theory, concentrating on what is meant by "sense". We postpone the issue of *how* senses might be generated until Chapter 4.

In the final section of this chapter, we turn away from the issue of senses and to the issue of word meaning. Traditional approaches to word meaning have rested on two assumptions. One, the Unitary Assumption, is that unambiguous words have one and

Contextual Expression	Example
Indirect Nouns	*one water, a Henry Moore*
Compound nouns	*finger cup, apple-juice chair*
Possessives	*John's dog, my tree*
Denominal nouns	*a waller, a cupper*
Denominal verbs	*to farewell, to Houdini*
Eponymous verbs	*to do a Napoleon, to do a Nixon*
Pro-act verbs	*to do the lawn, to do the porch*
Denominal adjectives	*Churchillian, Shavian*
Non-predicating adjectives	*atomic, manual*
Eponymous adjectives	*very San Francisco, very Picasso*

Table 3.1: Ten Types of Contextual Expression

only one meaning. The other is that such meanings can only be properly described in terms of necessary and sufficient conditions on word use. I will argue that the picture of sense which we have established requires a theory of word meaning to reject these two assumptions. In essence this calls for a revised notion of analyticity, which at this stage we will only crudely characterise. However, in Chapter 5 we will examine the role that the traditional conception of analyticity has played in arguments concerning proper names and how certain of these arguments may be circumvented were our revised notion of analyticity accepted. This, then, will pave the way for our discussion of natural kind terms in Chapter 6.

3.1 Motivating Sense Generation: Some Alternatives

Clark (1983) espouses a view according to which the senses of certain expressions, so-called contextual expressions, are not selected from some exhaustive list but are, rather, created on and for the nonce. Hence the title of his paper, *Making Sense of Nonce Sense*. In motivating his view, Clark begins by outlining various sorts of contextual expression and Table 3.1, reproduced from the original article, summarises this.

The problem with such expressions is that their senses are highly dependent on context.

	Alterability of Aspect of Meaning	
Aspect of Meaning	Fixed	Shifting
Sense	Purely Intentional Expression (e.g., *bachelor*)	Contextual Expression (e.g., *to teapot*)
Reference	Proper Name (e.g., *George Washington*)	Indexical Expression (e.g., *he*)

Table 3.2: Classification of Expressions

Take *apple-juice chair*, for example. Out of context, we may find it difficult even to imagine a possible sense for this expression. But in context, it is easy. If, for example, in a dining room there is in front of each place-setting a glass of different flavoured juice, then *apple-juice chair* can serve to identify a particular chair. That is, it may have the sense "chair in front of which is some apple-juice". Similarly, with the other types of contextual expression. *One water*, for example, may, depending on context, have the senses "one glass of water", "one bucket of water", "one drop of water", and so on. As for denominal verbs, Clark offers the following, by now well-known example.

> Imagine that Ed and I have a mutual friend named Max , who has the occasional urge to sneak up behind people and stroke the back of their legs with a teapot. One day Ed tells me, "Well, this time Max has gone too far. He tried to teapot a policeman".
>
> *(Clark, 1983; p. 301)*

Here, *to teapot* is being used with the sense "to rub the back of someone's leg with a teapot". We can construct other scenarios to further exemplify the properties of these expressions, but the point is surely clear. There can be no limit to the number of explanatory scenarios with which we can furnish a particular contextual expression and so, in principle, there can be no limit to the number of senses of said expressions. This is what Clark calls "non-denumerability". Further, the fact that the sense of contextual expression are so dependent on the context in which they are used suggests that said expressions are akin to indexicals. This is what Clark means by "contextuality". Indeed, Clark offers the classification of expressions reproduced in Table 3.2.

Now we may well quibble with Clark's classification, in fact we will do just this a little later in this section, but the classification serves the purpose for which it was intended, namely to highlight the fact that there are numerous expressions whose senses *shift.* They shift according to context and in ways such as to make the totality of senses non-denumerable. This, then, is the problem for standard views of parsing. For, as Clark indicates, most parsers assume that the senses of expressions are stored in a finite lexicon and that parsing consists, in part, of the selection of the appropriate sense from among these. This is what Clark calls the *sense selection* assumption. However, not only is it impossible to enumerate in a finite lexicon a non-denumerable set of lexical entries, it is also impossible to select from that set. The selection of a sense from a lexicon presupposes the search of that lexicon, and the search of a non-finite lexicon could not terminate. Thus, the sense-selection assumption is invalid, at least for contextual expressions. Though Clark, in his original article, appears agnostic as to whether the assumption is valid for other expressions, in the next section we will attempt to put this to the test.

Now, although Clark's concerns and ours are clearly related there is an important difference. Clark's paper explicitly addresses the problem for parsing created by the fact that certain expressions, contextual expressions, have non-denumerable senses. The difference lies in the fact that Clark explicitly excludes the conventional uses of words from this category of expressions. That is, "lemon" when used to refer to things which most certainly are lemons is deemed *not* to be a contextual expression. Clark discusses the case of *water.*

> *The water* could be used in the conventional sense "the substance called water" or in some nonce sense "the glass, or pail, or drop, or the teaspoon, or..., of water.
>
> *(Clark, 1983; p. 302)*

That is, for Clark, conventional senses and nonce senses are distinct. Indeed, any one of a number of quotes could make the point. Earlier, as we have seen, Clark classified expressions such as *bachelor* and *blue* as "purely intentional expressions", expressions

whose senses are fixed. Contextual expressions are expressions whose senses can shift. Later, Clark adds,

> If teapot were actually in the lexicon as a verb with the sense "rub the back of the leg with a teapot", then...I wouldn't have had to go beyond the conventional meaning listed in the lexicon.
>
> *(Clark, 1983; p. 326)*

Now it seems the point is clear. Conventional senses, such as the sense, "substance called water", of *water*, are indeed listed in the lexicon. Clark's conclusions concerning parsing stem from a consideration of contextual expressions and the implication is that considering the senses of conventional expressions does not force the same conclusion. Indeed, for conventional senses the *sense-selection* assumption holds: these can be listed in a finite lexicon. Clark & Gerrig (1983) are more explicit. A pure case of sense selection they give is the example of the senses associated with *radish*. Dictionary entries such as "plant of the genus *Raphanus*" and "pungent root of such a plant" are instances of conventional senses, senses which are finite in number and listed in the lexicon. It is on precisely this issue that Sense Creation and Sense Generation differ. As we shall see, one of the main motivations behind Sense Generation is precisely that what we consider to be the conventional uses of words belie many different senses, just, in fact, as with contextual expressions.

Arguing that conventional uses of words are like contextual expressions cuts across the classification schema in Table 3.2. Indeed, we may well quibble with other aspects of Clark's classification, though it is fair to say that the classification is not solely Clark's. Indeed, it is one which is standardly assumed in many different approaches to semantics. But, returning to the classification we are given, not only may we wish to assimilate many of his "purely intentional expressions" to the category of contextual expressions, but one might argue that proper names and pronouns could also be so assimilated. Consider the pronoun *she* and the degree to which we may maintain that it has a fixed sense. Now, indexical expressions such as these are generally thought not to have a fixed sense which determines their reference, rather their reference is determined by factors determined by the situation of utterance. However, one might be tempted to

suggest that one property that the referent of *she* must satisfy, regardless of the context of utterance, is that of being female. That is, one may suggest this property as a fixed aspect of the meaning of *she*. The "shifting" aspect of the meaning having to do with *which* female it is that is referred to in the context of utterance. However, even such a claim as this, seems to encounter difficulties in explaining instances where something is palpably not female, yet nonetheless being referred to as *she*. Male drag artists, for example, can, with great felicity, be referred to by uses of *she* and *her*. Similarly, in the case of individuals who have changed sex. In many instances, particularly for legal concerns, they are deemed not to have changed sex. Thus, if they were originally female, in the eyes of the law they are female still, and the word *woman* is taken to apply to them. However, in many other circumstances they are rightly described as *men*. That is, depending on circumstance, a word such as *man* may be taken to either apply or not to apply to one and the same individual. So to presume that words such as these and the related pronouns carry the fixed imputation that their referents are of one particular sex is a mistake. Rather, the imputation changes according to the circumstances in a way that likens them to contextual expressions. That is, while it may not be clear just precisely how these examples are to be accommodated, it is also clear that they cannot be excluded from the category of contextual expressions *a priori*. Rather the classification of pronouns as expressions whose sense is fixed is one that has to be argued for rather than stated by *fiat*.

Similar points can be made against Clark's classification of proper names. His discussion seems to assume that these are just as Kripke (1972) describes: *rigid designators*, expressions whose reference is fixed and whose sense, in as much as they can be said to have any sense, is also fixed. In Chapter 5, we will discuss the possibility that there is an alternative to this characterisation of proper names based on a revised notion of analyticity which Sense Generation seems to require. We postpone further discussion until then.

In the following discussion of theories of sense we will be concerned mainly with the apparently conventional uses of words and their associated senses and their assimilation to the category of contextual expressions. Consequently, though our arguments will be

related to those of Clark, there will be differences. We begin by introducing a little terminological difference: what Clark characterises by the *sense selection* assumption, we call Strong Sense Selection. We call this "strong" simply because, as a psychological theory, it makes stronger claims regarding the number of mental representations needed to account for the diversity of senses. We then contrast this with what seems a more plausible alternative, Weak Sense Selection. We conclude that neither are satisfactory.

3.1.1 Strong Sense Selection

Strong Sense Selection is characterised by Clark's *sense selection* assumption. Namely, that each and every sense of a given word is listed in the lexicon under the entry for that word and that what is required for successful communication is selection of the correct one in the circumstances in which the word is used. That is, in terms of the choices of Chapter 2, for each sense of an unambiguous word there will be a separate LEXON such that the descriptions comprising each sense and its corresponding LEXON are the same. Before we turn to some of the problems which this view faces, let us consider some of its implications for a couple of examples.

Turkey is a word which can be used with different intended senses. One can, for example, use it with the sense of "living, feathered, large, galinaceous bird" when talking of farmyard turkeys. Just as with Clark's denominal verb *to teapot*, we can picture a scenario in which this is the case. For example, in "The farm-hand fed the turkey", we would clearly want the sense of *turkey* to include the fact of their being alive, at least at the time of feeding. Another sense for *turkey* is one which expresses the fact that the animal has been killed and prepared for cooking. At Christmas, for example, when we ask if we are having a turkey for dinner, we are not asking if we are about to be served a living, feathered, large, galinaceous bird. These observations are problematic for a theory of sense which excludes conventional senses, for what is clear is that both farmyard and dinner-table turkeys are very definitely turkeys. It is not that some entirely novel sense of *turkey* has been created. Rather there are two related but different senses for the same word.

A more familiar example comes from the work of Betsy Macken. In her forthcoming paper, *Out of the Mouths of Babes*, she reproduces a number of children's conversations and one, given below, strongly suggests that the participants are attaching different senses to the word *father*.

> Hasan: My father...
> Richard: You don't got no father.
> Cassandra: He got a father.
> Richard: Not his real father.
> Cassandra: He got a father that lives with him.
> Hasan: My father gonna buy me a bike.
>
> *(Macken, forthcoming)*

As Macken points out, Richard seems to be aware of more than one "definition" for *father*: his first utterance seems to rest on the sense for *father*, "real father"; his second utterance relies on the fact that he knows *father* may have the sense "live-in father". In this case, of course, one may argue that only one of these senses should be labelled *conventional*. However, it would not be difficult to find an example where it seems that *father* had two senses both applying to conventional fathers. For instance, imagine a game involving fathers and their offspring where each father is paired with a child. Further suppose that one of the offspring, X, has reached a certain age and has become a parent. When an instruction is issued, as part of the game, to the fathers, the sense of *father* would be such as to exclude X, yet apply to all the other fathers. Even were all these fathers to be "real fathers", the example would stand: there would be two senses attached to *father*, one applying only to a subset of real fathers. A related example is where a sense of *father* applies to some, though not all, biological fathers and some adult males who may have no offspring. An illustration comes from the case of an undoubtedly real father, that is a *biological* father, who nonetheless is so neglectful and uncaring that neighbours claim that he is not a *father*. Indeed, they may exclaim that he is not a *real father*, while an adult male care-giver could well be described as a *real father* by such concerned folk. Hence, there is a sense of *father* which applies to some biological fathers and not to others.

Now Strong Sense Selection has only one way of treating these puzzles. The assumption

is that each and every sense is to be listed in the lexicon. So Strong Sense Selection supposes that there must be two lexical entries for *turkey* and *father*. In fact, it assumes that there must be at least two, since we cannot exclude the possibility of finding more senses for these words. So we will have two senses for *turkey* listed in the lexicon: one being "large, living, feathered, galinaceous bird", the other being "large, dead, plucked and cooked, galinaceous bird". For *father*, too, we will have at least two entries in the lexicon: one for the sense "biological male parent", one for "adult male care-giver". Though, if our suggestion above is correct, the number of entries will be far greater. This, then is the analysis, now let us turn to some of its problems.

One problem stems from the fact that there are important differences between senses over and above the fact that they may apply to different objects. One such difference is the distinction that may be drawn between those senses that are attached to a word by default and those that are attached only exceptionally. In the case of *father*, above, it seems that the sense "biological father" assumes the status of a default while the sense "live-in father" is more of an exception. The *sense selection* assumption holds that these different senses are simply listed in the lexicon: we are left none the wiser as to whether the distinction between default and exceptional senses is respected. Indeed as we have characterised Strong Sense Selection, this distinction is not respected. This is not to say that it could not be so respected, simply that the apparatus of Strong Sense Selection alone does not achieve this.

Now even were Strong Sense Selection to distinguish, somehow, between default and exceptional senses, there would still be an important deficit to its treatment. Namely, its assimilation of vagueness or generality to ambiguity. The issue, then, is whether these senses of *father* and *turkey* betray ambiguity. Let us return to some of the examples we have already considered.

Cruse (1986) offers some criteria by which one might establish that a given word is truly ambiguous. He enumerates both indirect and direct tests for ambiguity. For example, one indirect test is that the two senses of an ambiguous word may have different synonyms that, themselves are not synonymous. Consider, the case of *bank* exemplified

by the following pairs.

The bank is grassy.

The slope is grassy

The bank is wealthy.

The financial institution is wealthy.

Though *bank* and *slope* and *bank* and *financial institution* are not synonyms, they are close enough in meaning for the point to be made. That is, the two senses of *bank* seem to be similar in meaning to *slope* and *financial institution*, respectively. Further, since these are themselves not similar in meaning, we have some evidence for the claim that *bank* is ambiguous. That is, evidence for the independence of the different senses of *bank*. Cruse gives other indirect tests. The different senses of ambiguous terms should have different opposites or antonyms. Consider *light*.

The feather is light.

The feather is heavy.

The room is light.

The room is dark.

Here, the first two pairs show the antonymous relation between *light* and *heavy*, while the second two show the antonymous relation between *light* and *dark*. Since *heavy* and *dark* appear neither to stand in an antonymous nor a synonymous relation, this is further evidence of the ambiguity of *light*. We will not enumerate all of Cruse's criteria, but the addition of one direct test will suffice. He suggests the test of *zeugma*. When the different senses of an ambiguous word are simultaneously made active, a tension can result: this tension is labelled zeugma. The following serves to illustrate.

Jon and his bath continued to run.

Here *run* has two senses involving two denoted actions: one where a man is training for a marathon, one where the tap in his bath is gushing water. The strange quality of the sentence is attributed to the fact that for the correct interpretation of the coordinated sentence both senses of *run* must be accessed and that these two senses are antagonistic. That is, the simultaneous access of more than one sense of an ambiguous word is generally disallowed. This, then, is taken as further evidence of ambiguity.

Now none of these tests are infallible nor, even, some proportion taken together. Indeed, we could hardly expect it to be otherwise. Quine (1960), in his discussion of the ambiguity of terms (§27), indicates some of the problems of distinguishing ambiguous terms. While one line of evidence that a term is ambiguous, is the fact that it "can be clearly true or clearly false of one and the same thing" (§27, p. 131), he continues, "this trait, if not a necessary condition of ambiguity of a term, is at any rate the nearest we have come to a clear condition of it." So, perhaps definitive tests for ambiguity are not what we should expect. However, that we have some indicators is what we require for our continued exposition of Strong Sense Selection.

The question that should concern us is whether items such as *father* and *turkey* are indeed ambiguous, for this is how Strong Sense Selection treats them. By Quine's criterion they most surely are. We can envisage a particular individual of whom it can be said that he is and is not a *father*. Similarly, an individual turkey can, depending on the circumstances, be said both to be and not be a *turkey*. Cruse's criteria suggest, however, that, on the contrary, these words are *not* ambiguous. Take the test of synonyms, for example. It seems difficult, if not impossible, to find a synonym for a live, farmyard turkey that is not also synonymous with dinner-table turkey. Galinaceous bird, for example, will not do. And *dinner-table turkey* is, of course, not truly synonymous with turkey. Similar difficulties are encountered with the test of antonyms. The case of *father* is similar. What antonyms or synonyms of the two senses "biological father", "live-in father", could there be that would not themselves be synonyms?

The third test of zeugma, perhaps, provides the strongest evidence against the suggestion that these words are ambiguous. Consider the following.

Jon both fed and ate the turkey.

Here, there does not appear to be the same zeugmatic tension seen earlier. And indeed, taken together, the evidence is suggestive that *father* and *turkey* are *not* ambiguous.

It is useful to bear in mind that a claim of ambiguity amounts to two quite separate claims. Firstly, there is a claim about content. Secondly, there is a representational claim. With regard to the independence of these claims, we can see that we could, were we so inclined, suggest that the same content be repeated n times in a lexicon. This results in ambiguity, in a technical sense, but where such ambiguity is not manifest in any contentful way. Consequently, it is not surprising that the criteria outlined above are not in agreement. Quine's criterion, for example, is *prima facie* a claim about content. Can the same word be associated with mutually exclusive contents, is the question whose answer this criterion requires. But this is not, at first glance, a claim about representation. The tests of synonymy and antonymy are similar. They rely on sameness and oppositeness of content. and not on representational issues. The test of zeugma, however, relies not on some simple measure of content. It also appears to depend on the ability to access this content. That is, it depends, in some way, on the nature of the representation of the different senses of an ambiguous term. For Strong Sense Selection assumes that not only may there be different content associated with the same word but that these different contents are independently represented.

We can also add that the different senses of *turkey* and, perhaps to a lesser extent, *father*, are related. The different senses of *turkey* apply to individuals at different stages in some cycle, sadly, in this case, the life-cycle. The different senses of *father* apply to a disparate group of individuals who all bear some similar important relation to at least one member of the younger generation. The precise nature of the relation differs, but each relation seems similar to every other. An adoptive father is a father who assumes the conventional father role. A neglectful father is one who may be genetically related

to his charges but doesn't possess the loving attributes of conventional fathers. Treating these different senses as if they result from ambiguity is to ignore the important relations between them. Indeed, one of the characteristics of indisputably ambiguous words, such as *light* and *bank*, is the fact that their different senses are not so obviously related. The examples we have looked at so far have not, though, been of this kind and consequently the Strong Sense Selection treatment of these is untenable.

Our discussion so far has, then, suggested the following. Firstly, that there are different senses of words such as *father* and *turkey*. Secondly, that, for each of these words, the different senses are not independent. That is, the words are not strictly ambiguous. Yet this is precisely what is assumed by Strong Sense Selection. It assumes that the different senses of *father* are as independent as the different senses of *light*. This, then, is one problem with Strong Sense Selection.

Another problem relates to Strong Sense Selection's assumption that the number of senses of a word can be given in a finite list. Clark, as we have seen, is quite emphatic about this issue: contextual expressions do not possess a finite number of senses. So, with regard to contextual expressions of the sort detailed in Table 3.1, the assumption of a finite number of senses is clearly invalid. However, we have been concerned more with what might be called the conventional senses of Clark's "purely intentional expressions". The issue, then, is whether the number of senses associated with *father* in describing conventional fathers is finite.

This issue is difficult to resolve in a principled and unequivocal manner. Indeed, it is debatable as to what one would want to count as evidence either way. However, the example involving the game seems to suggest that *father* can indeed have an infinite number of senses. For the sense that *father* has in the circumstances is precisely determined by those circumstances, just as with Clark's denominal verb *to teapot*. The fact is that we can envisage infinitely many different circumstances in each of which *father* possesses a different sense. One of Mary Shelley's creations, Baron Frankenstein, could be said to be the *father* of another, the monster. An amoeba could also be said to be both *father* and *mother* to the two cells which result from its binary fission. While this

is a very different from the sense normally associated with *father*, it is clear that it is not unrelated. Nor would it be easy to argue that such a use of *father* amounted to a completely unconventional use as in the case of *to teapot*. Considerations such as these, suggest that there may indeed be no bound to the number of senses conventionally attached to a word such as *father*.

Perhaps the case of *turkey* is even clearer. The two senses we isolated earlier apply to the same individuals at two points in a process. One is live, the other has been prepared for eating. Yet there surely is nothing intrinsic to these stages in the process which allow *turkey* to have these senses, other than the fact that it is widely useful for us to distinguish them. The implication is that were we to find it useful to distinguish between other types of turkey, we could use *turkey* with different senses that would effect this distinction. Suppose, for example, that we wanted to distinguish between healthy and unhealthy turkeys. One person, suppose the one gathering healthy turkeys, may shout to a farm-hand standing in the turkey enclosure, *Do you have any more turkeys?*. *Turkey*, here, seems to have the sense of "healthy turkey" and only if the farm-hand were also a pedant would this have to be made explicit. Further, it seems that were the farm-hand to be a pedant and produce, in response, an unhealthy turkey, a reply of "That's not what I meant" could be expected. Such examples seem unlimited in number and this further suggests the unlimited number of senses that can be associated with a word.

Cruse is even more explicit on this point. In discussing the issue of the number of possible senses associated with words he has the following to say.

> One of the basic problems of lexical semantics is the multiplicity of semantic uses of a single word form (without grammatical difference). There seems little doubt that such variation is the rule rather than the exception: the meaning of any word form is in some sense different in every distinct context in which it occurs.
>
> *(Cruse, 1986; p. 51)*

Taking Cruse's point seriously necessarily commits one to the view that there may be infinitely many senses one might call conventional. So, perhaps our examples were not

necessary to drive home the point. However, the point should at least be clear. Strong Sense Selection assumes that all the senses of every word can be listed in a lexicon of finite size. All the evidence points to the fact that this assumption is quite without foundation.

Having found Strong Sense Selection theories wanting, we now turn to an alternative class of theories of sense, so-called Weak Sense Selection.

3.1.2 Weak Sense Selection

Weak Sense Selection is characterised by the view that many of the senses of words simply convey more specific information than that found in their corresponding lexical entries. So, what is required for successful communication is an initial selection from among some finite number of lexical entries possibly followed by some further specification of that lexical entry's content. That is, in terms of the choices of Chapter 2, for each LEXON there will correspond several senses, the descriptions that comprise the LEXON subsuming these corresponding senses. That is, the relation between a sense and the LEXON from which it comes is a monotonic one. Again, before we turn to some of the problems one might envisage for this view, we will consider some examples. The particular analysis we choose will not really be at issue: rather, the kind of analysis. We will be at pains to make this clear. Motivation for the Weak Sense Selection analyses will, further, rely on the discussion of particular examples found in the previous section.

Consider, again, the example of *turkey*. The Strong Sense Selection analysis required at least two lexical entries corresponding to its different senses. Weak Sense Selection, however, requires that whatever lexical entries we hypothesise, senses can be computed as a result of the specification of their content. If, for example, we hypothesise a lexical entry to correspond to the sense "large, galinaceous bird" then the two senses of the previous section can be computed by adding further information. The sense "farmyard turkey" may be added by adding the information that the bird is alive; that of "dinner-table turkey" may be obtained by adding the information that the bird is dead, plucked

and otherwise prepared for eating. The actual process by which information such as this may be added is presumably complex, but we can envisage it to be highly context dependent. So, senses are derived, in context, from a general lexical entry, a lexical entry which describes all entities to which the word applies.

Weak Sense Selection suggests that the lexical entry for *turkey*, for example, may apply to all entities to which the word applies, i.e., turkeys. Of course, many senses of *turkey* will apply to specific kinds of turkey, but the lexical entry must be sufficiently general so as to apply to all. There are a number of problems with this view which relate to the problems identified by Kripke (1972) and Putnam (1975). We postpone further discussion of the Kripke-Putnam view until Chapters 5 and 6. Here we will little more than hint at these problems.

One problem, though, relates to the fact that *turkey* should not apply to objects which are not turkeys. This places a requirement that whatever lexical entry we hypothesise it should not admit of entities to which the word does not apply. Another problem is that it should clearly apply to all turkeys. These, then, are the requirements that our lexical entry should specify necessary and sufficient conditions on the application of a word. That is, the conditions which are sufficient for the application of a word and the conditions which are necessary for the application of that word. The problem is that such conditions are notoriously difficult to specify as the arguments of Kripke and Putnam testify. Indeed, their arguments may convince us that such conditions simply cannot be found. This being the case for conventional uses, how much more difficult then must the problem be when we consider novel uses such as those that form the focus of Clark's attention.

Were there to be one general lexical entry, then, to account for all the conventional uses of a word, Weak Sense Selection would suffer major drawbacks. It would be difficult in the extreme to specify the content of such a lexical entry. One way of avoiding this pitfall is to hypothesise multiple lexical entries to account for all the conventional uses of a word. This, it seems, is the only plausible option for Weak Sense Selection.[1]

[1]One may argue here that this rather depends on the nature of the lexical entry. Prototype theory, for example, may be thought to circumvent this problem. Indeed, so might any theory where lexical

An example of this kind of Weak Sense Selection is provided by Cruse (1986). He considers the case of *dog* and two of its senses: "canines" and "male canines". The conjecture is that the second sense may be computed on the basis of some contextual influence. Were this to be the case, it would illustrate Weak Sense Selection perfectly. We could hypothesise the content of the lexical entry for *dog* to be "canines" so that the sense "male canines" may be derived from it by the adding of information supplied, perhaps, by context. In fact, *dog* seems more likely to be ambiguous but the point is hopefully made. Weak Sense Selection could, returning to the *turkey* example, hypothesise two lexical entries for *turkey* corresponding to the senses "farm-yard turkey" and "dinner-table turkey". The other senses we encountered, "healthy farmyard turkey" for example, can then be computed, again by adding information. Before turning to some of the problems of this view, let us see how the *father* example could be treated.

Father, under Weak Sense Selection, would be accorded more than one lexical entry (to avoid the Generality option) but less than the total number of senses. For example, it could have a lexical entry corresponding to "biological father" and one corresponding to "male adult assuming the role of father". Now, an adoptive father falls under the latter description but is not accurately described by it. Yet it seems clear that there is indeed a sense of *father* which can apply to just adoptive fathers. We need simply construct a scenario as before. According to Weak Sense Selection such a sense can be computed by the addition of information, in this case information pertaining to the facts of adoption. Similarly, the lexical entry can be used to compute a sense "foster father" for *father*. So, Weak Sense Selection can, with some assumptions about context and the supply of information, accommodate most of the puzzles we have discussed without recourse to a non-finite lexicon. However, there are problems.

Firstly, we have already noted the problem of specifying necessary and sufficient conditions. This appears to rule out what we have called the Generality option, the option of specifying one general lexical entry. A more realistic alternative is to consider specifying a finite number of lexical entries, though more than one, from which additional senses

entries are identified with cluster concepts. We will argue in Chapter 7, in particular, that such a move towards cluster-concepts is beset by other difficulties.

may be derived. The problems with this option though are, principally, twofold. One concerns the fact that it leads us to treat generality as ambiguity. The other concerns the fact that we may be ignoring an important psychological aspect of word meaning by treating related senses as independent. Let us take each in turn.

Let us consider the way we envisaged Weak Sense Selection treating the *father* case. One possibility was that we could have two lexical entries: one corresponding to the sense "biological father", one corresponding to the sense "live-in father". Were these the only two senses for *father*, we would have a Strong Sense Selection analysis. Accordingly, we might envisage that our objections to Strong Sense Selection could well apply here. Indeed they do. One major concern with Strong Sense Selection was that it postulated multiple lexical entries for words that are not seemingly ambiguous. The same is true of Weak Sense Selection. It is not as pernicious: not as many lexical entries are hypothesised. But, for all that, it treats *father* as an ambiguous word. We will not detail the arguments against this treatment as they can be found in the previous section. Rather, we turn our attention to the other problem with this option, that of ignoring the psychological factors relating to senses.

Much of the following discussion will trade on the psychological factors which underly the application of a concept to an object. Though this issue is clearly related to that of how a word may apply to an object, a discussion of this relation and its relevance to theories of sense is in order. We begin by considering some views from the psychological literature itself.

Throughout the psychological literature pertaining to concepts there are strong connections drawn with the psychology of word meaning. Indeed, often the assumption is made that word meanings and concepts are one and the same. This is an identification with which we do not agree but when we come to talk of senses in the next section we will say more about how these relate to concepts. For now, however, we will be concerned with conceptions that already pervade the psychological literature.

Numerous researchers seem committed to the close ties between concepts and word

meaning. Cohen & Murphy (1984), for example, are quite explicit in connecting theories of concepts to theories of semantics. They suggest that theories of concepts can be traced, via the model-theoretic tradition, to Frege himself. Further, they suggest that concepts have been viewed as definitions of terms (of some language) and that an appropriate question for a theory of concepts is how such definitions operate in language comprehension. Both of these suggestions indicate very direct parallels between the notion of sense, which can indeed be traced to Frege, and the psychological notion of concept. Indeed, the suggestion is that senses are none other than concepts.

Similar connections are drawn by Johnson-Laird (1987). In his discussion of meanings and prototypes (pp. 203–4) Johnson-Laird claims that the lexical entries for natural kind terms are likely to include information for identifying and imagining exemplars as well as other conceptual information. His discussion suggests that prototypes (a notion ostensibly developed as part of a theory of concepts) should be represented in the mental lexicon. That is, we are led to believe that prototypes play a crucial role in natural language understanding. For our purposes, then, we would need to consider how the various senses of a word are related to the prototype associated with them. That is, how senses associated with word use relate to concepts which form the stable mental representations underlying word use.

Connections between concepts and what we call senses can be found frequently in the psychological literature. Carey (1988), for example, uses the terms "concept x" and "meaning of the term 'x"' interchangeably to refer to mentally represented concepts and meanings. This is justified, she adds, by the fact that supposed differences between child and adult mental lexical entries always corresponded to a conceptual difference. Again, the suggestion is that whatever comprises an entry in the mental lexicon, it is a concept.

What I have tried to demonstrate here, is that, for a psychological input into our theories of sense, we need to examine psychological theories of concepts. Much of this will be done in Chapter 7. Here we concentrate on a particularly influential argument regarding the coherence of concepts, an argument that suggests the mistake in treating genuinely

related contents as independent. It is here, in the treatment of genuinely related senses as independent lexical entries, that we will find Weak Sense Selection lacking.

Categories cohere. That is, there is *something* that brings objects together to form categories and whatever concepts are it is this "something" that they should specify. In exploring the arguments concerning conceptual coherence and their implications for Weak Sense Selection we will consider a number of examples, though we will concentrate later on a possible analysis of *mother*. A number of consequences follow from this treatment and, in particular, we will argue that we cannot maintain both Weak Sense Selection and a theory-based view of coherence. To see this we must turn to Murphy & Medin (1985).

Murphy & Medin (1985) distinguish between internal and external components of conceptual coherence. Both aspects derive from theories of the world. Internal coherence results from the relations between properties (described, perhaps, in terms of attribute-value pairs) for which the theory provides an explanation. For example, the various properties of chairs represented in the concept (its size, having a seat, its rigidity and so on) cohere in virtue of theories concerning the function of a chair. External relations between different concepts are also motivated by theories that support causal and explanatory connections between those concepts. Consider the concepts *cat* and *dog*. According to Murphy & Medin these concepts are coherent to the extent that they are embedded in our theories of the world: the greater the number of connections between such concepts and the rest of our knowledge, the more coherent and stable they are. The connection of *cat* to many different aspects of our knowledge (such as our theories of domestic life, ownership, pleasure and so on) provides it with coherence over and above that which is provided by our theories of *cat* alone, which is the source of the concept's internal coherence. The same thing is true of complex concepts, for example *cat and dog*. The concept is coherent as a result of there being many common theories of cats and dogs. Seemingly, on Murphy & Medin's view, we could not allow that complex concepts such as *elephants and lemons* or *stone lions and trout* have the same level of coherence as *cat and dog*. This stems from the fact that such complex concepts have fewer common theories to relate the component concepts.

This view of the theories underlying coherence seems to be intimately connected to metaphysical issues. Accordingly, for Murphy and Medin, concepts result from the way that theories cut up the world. Consequently, categories that cut across ontological boundaries would not be very coherent. For example, *elephants and hopes* would not be considered coherent, unless this category were motivated by a theory. If this theory were consistent with one's other theories, the coherence of the concept would be enhanced. The claim of Murphy & Medin is that the flexibility of conceptual coherence derives from the flexibility of particular theories. On our interpretation of their claim, this means that there are two mechanisms for allowing coherence to be flexible. Either, in certain contexts, coherence is enhanced by constructing an explanation or theory or, alternatively, the theories that underpin coherence are inherently flexible. It is not clear how these mechanisms for flexibility in coherence would work for examples like the Lion puzzle in which there are two different, yet apparently equally coherent, senses associated with the same word, *lion*.

The Lion puzzle shows that the word *lion* can be used quite appropriately to talk of a statue of a lion. However, there is a question as to just how many of our theories of lions involve statues. Seemingly, our theories of lions are capable of cutting the world up so that we can categorically assert that a statue of a lion is not a lion. One approach to this phenomenon is to view it as a peculiarity of word use. However, we wish to preserve the intuition that there is an essential perspectival aspect to this case. That is, we can adopt a different perspective on a statue of a lion: we may simply view it as a lion. Additionally, in this case each of these perspectives seems equally coherent. As we see it, there are two possibilities for Murphy & Medin to account for this apparent flexibility in coherence. One is to invoke their suggestion regarding the flexible nature of theories. Another is to adopt the standpoint of Weak Sense Selection, and thus hold that there are two lexical concepts for *lion*. In what follows we will explore the former possibility and conclude that the latter is their only plausible option. This in turn leads to an inconsistency.

Under the first option, that of invoking flexibility in theories, we envisage three possible mechanisms open to Murphy & Medin for achieving the flexibility of coherence. One

mechanism that Murphy & Medin offer is that of the construction of an explanation in cases such as *elephants and hopes* where, seemingly, ontological boundaries are crossed. In the above case, apparently there is a similar crossing of ontological boundaries. Our theories of lions tell us that stone lions are not really lions, though presumably there are some theories of lions that are also theories of stone lions. Perhaps, then, it is that the coherence of the concept of lion in this case relies on a constructed explanation. However, since our stable, underlying theories concerning lions cannot all apply in the case of the stone lion, it must be that the use of lion in talking of a stone lion is less coherent than the use of lion in talking of a real lion. This follows because coherence in Murphy & Medin's view falls out of the number of explanatory links which structure the concept. Hence, on this view such constructed explanations necessarily lead to less coherent concepts than those stable explanations provided by the knowledge base. This is at odds with our intuitions. Seemingly both uses are, in this case, equally coherent. What we want is to allow that both ad hoc and stable explanations can provide for equal coherence.

Murphy & Medin provide another mechanism to account for the flexibility in coherence that we observe. This is the possibility that we may have flexible theories. In the case of the stone lion, we might suppose that our theories concerning lions are flexible, accounting for the observed flexibility of coherence in use. The use of *lion* to talk of a stone lion can be coherent as we would want in virtue of the fact that the same theories concerning lions can also be theories concerning stone lions. However, this does not accord with the intimate connection that theories have to metaphysical concerns. If our theories concerning lions are also theories concerning stone lions then we have no principled manner of cutting the world such as to differentiate between non-lions (like our stone lion) and real lions. We take it that this undermines the metaphysical position that Murphy & Medin adopt.

A third possibility that we believe Murphy & Medin may allow for is the option of theory change. This option entails that in order to preserve the coherence of the lion concept when employed to talk of a stone lion, the theories underlying the concept would have to change. This, however, seems problematic. Since they adopt the Quinian position

of viewing theories as components of a large and intricate web, a change in one will have ramifications for the others. In this case we would have to change our theories concerning lions so as to allow the possibility of inanimate lions. Changes as radical as this, as they acknowledge, would require a global reorganisation of one's knowledge base. It would also require, under most interpretations, a divorcing of theories from metaphysics. This is precisely what we argue for. In Chapter 4 we will present a view that is not tied to the assumptions that lead to these counter-intuitive consequences. As we have seen there are good reasons to suppose the kind of flexibility observed in coherence is not to be captured by the notion of flexibility in theories. We now turn to the second possibility we pointed to earlier.

The alternative position that we believe Murphy & Medin might adopt is the one that we call Weak Sense Selection. That is, the above arguments will not go through if we assume that there are two lexical entries and hence two lexical concepts for *lion*. Aside from the arguments already presented against this option, it seems there are further considerations that would prohibit Murphy & Medin from taking this view. Following Quine (1960), we might like to distinguish between two lexical entries for *light*. This move seems legitimate. After all, it seems that the theories we have for light-weight objects do not have to be the same as those for light-coloured objects. For the case of *lion*, although lions and stone lions are differentiated by the theories provided by the knowledge base, lions and stone lions are nonetheless very closely related. It is presumably the case then that since we can relate lions and stone lions we have theories to do just this.

Returning to the Weak Sense Selection of *mother*, we can see that similar arguments apply. Weak Sense Selection holds that words such as *mother* have multiple lexical entries, perhaps ones corresponding to the senses "biological mother" and "female adult care-giver". If we tie Murphy & Medin's arguments to this position then we can see the possibility that these different senses are indeed not independent, but related by theories. Were this to be the case, positing independent lexical entries would ignore the psychological import of such theories. Consider the following example.

> Kim lives with her biological mother, Mary. Mary adopts a child, Jane, to whom Mary behaves in all respects as she does to Kim. Jane calls Mary "mother" and behaves towards her in the same way Kim does. Initially, Kim cannot accept that Mary is Jane's mother but after some time she does.

How is it that Kim's notion of mother appears to change? One possibility is that initially she only has the *biological mother* concept of mother and that she somehow acquires another concept of mother, the *adoptive mother* concept. Even though Kim starts off with the notion of mothers as biological mothers, one can envisage that Kim's theories of the world allow her to construct an explanation of how Jane can call Mary "mother". Indeed, she may come up with a new notion of mother, a mother who is like a biological mother in all respects save the biological relations. That is, Kim, with her notion of a biological mother, equipped only with her theories of the world, can construct the notion of an adoptive mother. The fact that this seems so clearly to be the case undermines the starting point of Weak Sense Selection. Namely, that there may be independent lexical concepts for *mother*. The fact that they may be related by our theories of the world indicates that they are not independent. We should note here the fact that different words do have contents which may be related by theories. Yet the fact of their relation could not be suggested as undermining their existence as different words. So, our argument does not concern simply the fact that the contents of the senses of *mother* may be related, but that they are related in a certain way. In particular, what the example suggests is that the content which may be associated with one sense is derivative, in some way, of the content associated with another sense. For instance, the content of the "adoptive mother" sense of *mother* may be derivative on the content of the "biological mother" sense of *mother*.

We have seen, then, that to posit independent lexical entries is to undermine those very arguments that Murphy & Medin advance in favour of theories. Yet, we take those arguments as very good grounds to accept the theory-based account of concepts. Consequently, we also take it that what is needed to avoid the pitfalls of the sense selection accounts is an extension of this theory-based view. So, although we crucially need a way of describing the difference between lions and stone lions and mothers and adoptive mothers, we need a way of avoiding making an unprincipled distinction at the

level of lexical entries.

There are several possible responses to this difficulty. One is to place constraints on the way theories may relate these independent lexical entries. That is, we may want to place constraints on the possible extensions of any two of the lexical entries for *mother*, such that one cannot be extended to form the other. However, this flies in the face of the observation that we can do just this. Kim *can* construct a notion of adoptive mother from her notion of biological mothers. So for this to be a viable option, more justification is needed. Another option is simply to reject the arguments of Murphy & Medin that the coherence of senses is determined by theories: the fact that biological and adoptive mothers may form a coherent complex concept has nothing to do with the theories that relate them. As stated earlier we agree with Murphy & Medin's general position. We believe that the only other alternative is to reject the assumptions on which Weak Sense Selection is based.

A consequence of considerations such as those of Cruse (1986) and Murphy & Medin (1985) require theories of sense and word meaning to posit different lexical entries only on the basis that a word has senses which are genuinely unrelated in the sense indicated above: that is, the content of one sense must not be seen as being derivative in some way on the content of the other. For examples such as the ones we have examined above, we are committed, then, to the view that words should receive only one lexical entry. One suggestion we have already considered which would assign only one lexical entry to *father*, say, is the Generality option. In many respects, the Generality option is close to what we would expect of a theory of sense. It respects both the psychological and linguistic evidence we have considered. Where the Generality option singularly fails is in the fact that lexical entries may contain little or no content. The reason is clear. Since words can apply to so many disparate objects, there are really very few commonalities between them. Consequently, very little can meet the twin criteria of necessity and sufficiency for the conditions of application of a word. Consequently, the content of lexical entries must necessarily be sparse. Another failing of the Generality option is its seeming inability to distinguish between default and exceptional senses. Even the content of default senses must presumably be inherited just as with exceptional senses.

These unfortunate consequences result from one assumption underlying the mechanism by which senses may be derived from lexical entries. In our preceding discussion we have assumed that this mechanism is fundamentally monotonic. It is the rejection of this in favour of a mechanism which is fundamentally non-monotonic which is the hallmark of Sense Generation.

3.2 Characterising Sense Generation

The previous sections have been concerned to detail what Sense Generation is not. In this section we attempt to say what it is. The term *Sense Generation* presumably admits of a number of distinct theories of sense. We will, in Chapter 4, outline a particular version of Sense Generation called the Relational View. The purpose of this section is simply to convey, in an intuitive way, the properties of Sense Generation approaches, some of the theoretical apparatus on which they rest and some of the formal apparatus in terms of which such approaches can be described. We briefly consider some of the psychological claims that one may associate with such apparatus.

Sense Generation is a framework for the treatment not only of contextual expressions such as Clark has outlined, but also of the conventional senses of words. It attempts to marry linguistic considerations concerning ambiguity with psychological evidence concerning coherence. It also attempts to respect the distinctions between default and exceptional senses. As such it embodies the following principles. Only independent senses should derive from independent (at least, synchronically) lexical entries. A lexical entry for a word should contain the information associated with the default sense of that word. Non-default senses are, it is assumed, generated from the lexical entry by a process which is fundamentally non-monotonic. Importantly, the process by which these senses are generated is envisioned to be one that provides the sense with its degree of coherence. That is, in terms of the choices of Chapter 2, for each unambiguous word there is a single LEXON from which senses may be derived, their content possibly being related non-monotonically to that of the LEXON. We will discuss the integration of such a view of senses with a view of meaning in Chapters 5 and 6. For the present let us

concentrate on Sense Generation in itself.

The first consideration is the result of our previous discussion concerning linguistic evidence for ambiguity and psychological evidence concerning coherence. Our previous discussion has shown that given these considerations, we are unable to assign, say, *lion*, two different lexical entries: one corresponding to "real lions", one to "statues of lions". The second consideration indicates that the lexical entry for *lion* should embody that information associated with the default sense for *lion*. Borrowing the familiar notation of feature-structures, we might expect such a lexical entry to have the following content.

$$\begin{bmatrix} \text{ANIMATE}: & + \\ \text{FIERCE}: & + \\ \text{COLOUR}: & \textit{brown} \\ \text{LEGS}: & 4 \\ \ldots: & \end{bmatrix} \tag{3.1}$$

Whether these are the correct attributes and values to describe the default sense of *lion* is, presumably, a matter for empirical study. So we are not committed to, and nor does our account rely on, these particular attributes and values. Rather, we are using these particular attributes and values simply to demonstrate the nature of Sense Generation. The third and fourth considerations require that the sense of *lion*, "statue of lion", must be derived from the lexical entry described above. Such a sense might be as described below.

$$\begin{bmatrix} \text{ANIMATE}: & - \\ \text{FIERCE}: & - \\ \text{COLOUR}: & \textit{brown} \\ \text{LEGS}: & 4 \\ \ldots: & \end{bmatrix} \tag{3.2}$$

The derivation may be, and in this case is, non-monotonic. The mechanism for such a derivation must also be one which provides this sense with its coherence. In Chapter 4,

we will examine in more detail a proposal concerning the nature of this mechanism. Before we do so, however, we will say a little more about the kinds of object which play a role in Sense Generation. In particular, we will say more about the notion of sense we are employing.

3.2.1 Senses

We take *senses* to be descriptions of word meanings. Further, we assume that these must be both privately, mentally representable and publicly specifiable. It is important to note that our commitment to senses does not constitute a commitment to Frege's "third realm", distinct from the mental and physical realms of objects. The most important aspect of senses for our purposes is the way in which we assume that they guide linguistic behaviour. In particular, we assume that the application of a word to an object is mediated by the *sense* of that word. That is, we assume that the description which constitutes the sense is satisfied by that object. This assumption requires that the correct uses of a word must be explicable in terms of the sense or senses which that word possesses. As such, we can take senses to classify the linguistic behaviour of cognitive agents. An example will clarify what is meant.

Suppose our cognitive agent, Fred, is asked to identify lions in his immediate vicinity and suppose further that despite being aware of a nearby statue of a lion he replies that there are no lions in his immediate vicinity. Supposing that senses classify linguistic behaviour, we can posit that Fred's understanding of "lion" relied on a sense of "lion" which excludes statues of lions. We might, for example, posit a sense which applies only to *real* lions. However, on other occasions Fred can respond that there is a lion nearby. Such a use of "lion" here seems to depend on a sense for "lion" which includes statues of lions.

This view of senses is closely related to that espoused by Evans (1976). In his discussion of the Fregean notion of sense, Evans claims that underlying Frege's position is a commitment to what he calls the "intuitive criterion of difference" for thoughts.

"Thought" is the term Frege used to describe the senses of declarative sentences and, as is well-known, he viewed these thoughts as being composed out of the senses of the sentence's parts. Evans describes the intuitive criterion of difference for thoughts in the following way:

> ...the thought associated with one sentence S as its sense must be different from the thought associated with another sentence S' as *its* sense, if it is possible for someone to understand both sentences at a given time while coherently taking different attitudes towards them, i.e., accepting (rejecting) one while rejecting (accepting), or being agnostic about, the other.
>
> *(Evans, 1976; pp. 18–19)*

Evans cites the example Frege himself used to illustrate this intuitive criterion of identity. Frege describes the case of the mountain which has the two names Ateb and Aphla. Since it is perfectly possible for someone to understand[2] both the sentences "Ateb is at least 5000 metre high" and "Aphla is at least 5000 metres high" and yet accept one as true and the other as false (or at least to be agnostic about its truth), the sentences must have different senses. This despite the fact that the two sentences have the same truth value or Fregean Meaning. Given that the only difference between these two sentences is the proper name in subject position, it follows from the intuitive criterion of difference for thoughts and Frege's principle of compositionality that these proper names have different senses. Different words, then, even though they may have the same referent, can have different senses. However, this is somewhat different to the case of Fred described earlier. Here, if we are to explain the differences in linguistic behaviour that Fred exhibits towards the word *lion* in terms of senses, then we will have to demonstrate that this one word has different senses.

For the intuitive criterion of difference for thoughts to imply that one and the same word has different senses, we must consider one and the same rational agent and her understanding and attitude towards one and the same sentence at a given time. Consider Fred, contemplating the statue before him, being approached by both the art student and the zoo-keeper, and, this time, each uttering the sentence *That is a lion* (with

[2]Though one may, of course, argue that if there was such an individual then they could not possibly be said to *understand* these sentences.

suitable ostensive gesturing so as to leave Fred in no doubt as to the subject of the proposition(s) expressed). The issue is whether or not Fred can take different attitudes to these different uses of the very same sentence where the same entity is indicated by the demonstrative.

The issue seems to isolate a genuine choice point for theories of meaning and sense, related to those we isolated in the previous chapter. The position we adopt will almost certainly have significant ramifications for other aspects of our theory. For example, we could adopt the position that Fred does not (at least, coherently) adopt different attitudes towards the sentence *That is a lion.* However, such a position is one which prohibits the difference in linguistic behaviour which Fred exhibits towards this sentence from being explicable at the level of the attitudes, in particular, at the level of thought. Given our interest in theories of word meaning and theories of concepts, such a position is not one we will adopt. That is, our goal is one of explaining behavioural differences such as these in terms of the contents of thought and the contents of concepts. So, whereas some may argue that the use of *lion* by the art student is metonymic or metaphorical, that is, to be distinguished from conventional uses, we will not argue along these lines. Instead, we assume that Fred can coherently adopt different attitudes to this sentence and, hence, that, by Frege's criterion, *lion* has two senses at least.

So much for senses. In the last section of this chapter and in chapters 5 and 6, we will return to the claim, already mentioned, that senses are descriptions of meanings. In particular, we will argue that this view of senses commits us to an alternative to the traditional view of meaning, one which can be found, in essence, in Situation Theory. While this will form the semantical foundation of our view, for the present it suffices to note that we merely require that senses provide a psychologically plausible classification of linguistic behaviour.

3.2.2 Words

In tying senses to words, it is only proper to say a little more about words themselves. First, we should be clear what we mean by *word*. The problem here is that *word* can be used to mean a number of different things. As Lyons (1968) points out, and Brown & Miller (1980) reiterate, such a fact forces care in terminology.

One sense of *word* can be used to describe word *forms*. For example, in this sense, *laughed* (as in *I laughed*) and *laughed* (as in *I have laughed*) are instances of the same word. That is, the same phonological or orthographic string. The fact, though, that these different word forms have different syntactic properties (*laughed* in *I laughed* is in the simple past tense form, while in *I have laughed* it is in the past participle form) allows us to claim that they are really different words. Here, *word* has the sense of *morphosyntactic* word. Throughout this thesis, however, we will in general mean neither of these when we talk of words. Rather, we are more concerned with the sense of *word* meaning *lexemes*. Before we turn to LEXONs we will say a little more about this notion of lexeme in order to deflect a criticism of our approach.

In the analysis of Chapter 4, as we shall see, little is said about the diversity of senses attached to verbs. Indeed, the analysis trades almost entirely in common nouns. The reason is the difference that can be seen between nouns and verbs when one looks at their inflectional morphology.

The intuition that different word forms can be related to one another leads to the abstract notion of word called *lexical morpheme* or *lexeme*. Different word forms can be analysed morphemically as in the following example: *laugh* can be analysed as {LAUGH + Present}, *laughed* as {LAUGH + Past}. In addition, verbs may belong to various aspectual classes. *Run*, in *Jon ran* for example, may be assigned the *process* aspectual type. In *Jon ran a marathon*, *run* may be assigned a *culminated process* reading (cf. Moens & Steedman, 1988). Such observations as these provide difficulties for an approach concerned with the lexical semantics of verbs such as *run*. Consider, for example, the following pair: *Jon was running* and *The tap was running*. Clearly, these

pairs illustrate the fact that the verb *run* can be used to refer to different sorts of process, the problem is to decide on the locus for this difference. The intuitive criterion of difference requires not a pair such as this, but two readings deriving from one and the same sentence. In the case of Mt. Ateb we required sentences predicating the same concept of the same entity. Here we may do the same. *This is running*, for example, with suitable gesturing to the water pouring forth from the tap, may suffice, but the result is an unusual sentence to say the very least.

However, this is not the only option. We could just as well question the possibility of coherently adopting different attitudes to a sentence such as *The tap is running.* Indeed, it seems almost certain that one can. However, we still have a difficulty in deciding on the locus of the change in sense that the intuitive criterion of difference requires. Should we, for example, simply state that *running* has two different senses? And, if so, does this result from the fact that *run* can have those different senses or has it something to do with the simple progressive form? Let us take another example.

Compare *Jon demonstrates* with *Jon is demonstrating.* Although the case is far from clear cut, it seems that the first sentence has a primary reading in which *demonstrate* means to illustrate by example. Similarly, the second seems to have a primary reading in which *demonstrate* means to protest publicly. Were we convinced that the senses of *demonstrate* were generally different in these cases, then the problem comes in deciding the locus of this change in sense. Is it due to the verb stem *pace* lexeme, the copula or the aspectual class? With common nouns, fortunately, such problems are all but obviated.

Concentrating on common nouns, then, allows the locus of change in sense to be more easily decided. This is not to say that, in the case of morphologically more complex words such as verbs, deciding on the locus of change is beyond our means. Indeed, we assume that this is not the case, in principle, though it may be difficult in practice. Rather, our concentration on common nouns is in order that the exposition of Sense Generation is as clear as possible and that we are not distracted by issues concerning, for example, morphology, which strictly, are not Sense Generation's main concern.

3.2.3 Lexons

Throughout the literature on the mental representation of the meanings of words there is an assumption that words receive entries in a "mental lexicon". Just as for dictionaries, the entries of mental lexica include orthographic, phonological, morphological, syntactic and semantic information. Finding the meaning of a word is a question of "looking up" the appropriate entry in the mental lexicon and reading off the relevant semantic information (cf. Clark, 1983; 1989). Of course, as we have suggested in section 2.2.4, this metaphor does not commit one to any particular representational claim: it does commit one to a claim that such-and-such content is represented, but it does not commit one to any particular view as to how this content is represented. If, then, we are to correctly classify the linguistic behaviour of agents then our theory must allow for something to play an analogous role to such mental lexical entries. We choose to call the "semantic" aspects of such entities LEXONs, partly to avoid spurious and unwanted connotations of other more familiar terminology (cf. lexical concept). When we talk of LEXONs we mean, quite simply, those descriptions that we take to define the stable mental representation of the semantic contributions of words. And these descriptions are descriptions of worldly properties that we associate with the default uses of words.

Chapter 4

The Relational View

> Instead of producing something common to all that we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all, — but that they are *related* to one another in many different ways. And it is because of this relationship, or these relationships, that we call them all "language".
>
> *(Wittgenstein, PI, §65)*

In the last chapter, we spelt out, somewhat abstractly, a view by which the sense of a given use of a word, may be generated from that word's lexical entry. However, it is not difficult to see that, given our abstract characterisation, it seems that we are committed to the view that words may, in principle, have any sense whatever. In fact, unsurprisingly, such a characterisation requires a *caveat.* And it is the purpose of this chapter to spell out a particular view of Sense Generation, the Relational View, which provides just such a caveat. In particular, the Relational View is an attempt to provide some psychologically motivated constraints on the process of generating senses from LEXONs. Much of the psychological motivation comes from the literature on concepts and, as such, is best viewed in the context of extant theories of concepts. However, we will do this in Chapter 7, where we consider more fully the implications of the Relational View for a theory of concepts.

First, then, we briefly consider some of the psychological motivation behind the Relation View. Such motivation is best seen in terms of morals which derive from the existing

psychological literature. Having done this, we spell out the Relation View in some detail, attempting to account for some of the puzzles of Chapter 1 and, in the meanwhile, showing how the morals are observed. Our exposition of the Relational View will be in terms of situation theory and, in some instances, we will make reference to a prolog program which has been used to assist in the development and exposition of these ideas.

4.1 The Morals

In order that we may draw some morals concerning theories of word meaning and sense, we need to look at some theories that have been proposed. The discussion of these theories we leave until later (Chapter 6). And, indeed, a more detailed exposition of the arguments which suggest these morals will also appear later (Chapter 7). For the purposes of this chapter we simply need to consider the kinds of descriptions that have been offered in the psychological literature to explain the behaviour of words. The descriptions fall roughly into two types: those suggested by classical theory; and those suggested by prototype theory. Let us consider each type of description with respect to the word *lemon*.

Classical theory holds that the meaning of words and their corresponding concepts comprise a set of necessary and sufficient conditions, or a set of properties common to all the elements to which the word refers. The following exemplifies this claim with respect to the word *lemon* and the description can be taken as constitutive of the concept *lemon*.

$$\begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{oval} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix} \tag{4.1}$$

Prototype theory, also a theory of concepts, is committed to rather different descriptions as exemplifying the content of concepts and, hence, as determining the behaviour of

words. Indeed, the theory is one that adopts the position of explicating concepts in terms of a "cluster" of descriptions. As such there are weightings attached to attributes and also a specification of possible values for each attribute. There are more properties of prototype concepts, but we will leave this until later. The following, then, is the kind of description that might be associated with the word *lemon*.

$$\left[\begin{array}{lll} \text{ISA:} & \textit{fruit} & \\ \text{COLOUR:} & \left\{\begin{array}{ll} \textit{yellow} & 8 \\ \textit{green} & 3 \\ \textit{blue} & 1 \\ \ldots & \end{array}\right\} & \\ \text{SHAPE:} & \left\{\begin{array}{ll} \textit{oval} & 6 \\ \textit{flat} & 2 \\ \textit{round} & 1 \\ \ldots & \end{array}\right\} & \\ \text{TASTE:} & \left\{\begin{array}{ll} \textit{acidic} & 5 \\ \textit{sweet} & 1 \\ \textit{sour} & 1 \\ \ldots & \end{array}\right\} & \end{array}\right] \tag{4.2}$$

The numbers to the right of each value indicate the weighting of that particular attribute-value pair and the braces indicate that the values contained within are disjoint with respect to that attribute. Now, with these simple descriptions at our disposal we can outline some of the psychological considerations that we take to motivate the Relational View and which, in Chapter 7, we will consider in more detail.

Moral 1 *The Moral of Coherence*

Categories cohere. That is, there is *something* that brings objects together to form categories and whatever concepts are it is this "something" that they should specify. According to the classical theory of concepts, coherence is determined by attribute-value matching. That is, to determine whether a given object is a member of a category, we must determine whether the object satisfies the properties indicated by the attribute-value structure representing the concept. Assuming that the object may be described by

an attribute-value structure, then we would require this to be subsumed by the attribute-value structure corresponding to the concept. For prototype theory too, coherence is determined by a similar process of attribute-value matching. There are a number of arguments, though, why an attribute-value matching view of coherence is going to be insufficient to capture the full rich structure of real-world categories.

Recent discussion has led to the view that representing coherence is necessarily going to involve what might be described as "knowledge-rich" techniques. As a simple example we might consider how a malformed lemon that has become rotten and turned a peculiar shade of purple would still be categorised as a lemon. Seemingly, this can't be modelled in terms of similarity as computed by attribute-value matching since very few of the attribute-value pairs of either (1) or (2) would match those describing the object. Murphy & Medin (1985) and Medin & Wattenmaker (1987) argue that similarity computed by attribute-value matching is simply insufficient to determine the coherence of a category. Instead, Murphy & Medin (1985) propose that a major part of the coherence of a category is determined by "theories and knowledge of the real world" (p.312), these, in turn, being composed or "made up" (p.313) from concepts. Theories like those, for example, which express knowledge of such relations as flattening, painting, ripening, etc.

In trying to glean psychological truths for our account of word meaning this position seems circular. If we accept for sake of argument that senses can be described by concepts then Murphy & Medin's arguments suggest that senses are to be represented in terms of "theories" which in turn are to be represented in terms of concepts, these also describing senses. Murphy & Medin reject the suggestion that their position is circular, a clearer rejection can be made, I believe, once a particular assumption is abandoned: the Unitary assumption.

Underlying both the prototype and classical approaches is an assumption, the Unitary assumption, that a single, "basic-level" category (Rosch et. al., 1976) is to be represented by something corresponding to a single, unitary data structure, variously called a concept, lexical concept or, as we call it, a LEXON. It seems that it is this assumption

together with Murphy & Medin's arguments concerning coherence which appear to lead to circularity. The problem arises because a sense or concept is seen as both being a component of a theory and being composed of theories. For an account of word meaning, this suggests that underlying various uses of a single word is a single sense, itself being comprised of various other senses, these giving the single sense its coherence. So, the circularity emerges when we stipulate that, in fact, it is only one sense that underlies all the uses of a given word and it is this that leads us to think of senses as being composed of senses and the circularity that implies. We can avoid this circularity by denying that the various uses of a word all depend on a single sense. Rather, we may allow that these various uses may depend on a number of different, though related, senses. To conclude, any theory which attempts to represent all of the uses of a word in terms of a single sense can be ruled out by the moral of coherence *a priori*.

As an alternative, it seems that a much more intuitive account of conceptual coherence is forthcoming when we consider some more examples like the lemon. Intuitively, it seems that a dinner-table turkey is a turkey, not because of some *particular* set of properties that it has, but because we know of relations that hold in the world such that the properties of the object we observe (dinner-table turkey) stand in this relation with the properties of normal turkeys. Similarly, we can still talk of a dismantled bicycle as a bicycle apparently not because of some similarity as computed by attribute-value matching but because of our knowledge of dismantling operations. Murphy & Medin (1985) make a similar point. They state that "much of our reasoning about concepts may be based on constraints about operations that are permissible" (p.295). This point seems fundamental and we might expect that this clue could help in developing our account of word meaning. In particular, if some of the senses which underly the uses of a word express facts about such permissible "operations", then it seems our account will have more promise in respecting the psychological arguments regarding coherence.

Moral 2 *The Moral of Representational Economy*

In principle, there seem to be many ways of representing a category by encoding information in the attribute-value structures we use to describe senses. One way might

be to record, for every exemplar to which the word applies, the value of each attribute pertaining to the category. For example, if shape is relevant to the category of lemons then for each lemon we encounter, we could simply record its shape. This is the strategy that prototype theory adopts: if, equipped with (2) as our representation of the category of lemons, we encounter a flat lemon, the prototype structure will be revised. The disjoint set of values for the attribute "shape" will be expanded to include "flat" with a suitable weighting indicating its typicality. Such a strategy seems grossly inefficient and also fails to respect the following intuition: concepts are hypothesised to capture generalisations in order that finite brains may represent infinite possibilities. The strategy also seems implausible from the point of view of cognitive development, a point that has been made by others (for example, Cohen, 1983; p.87). As an alternative, we might consider that it is our knowledge of operations and other relations in the world that allow us to represent categories in the way that we do. Such knowledge should not be encoded in each representation we have for different classes of objects as it is in prototype theory: we wouldn't want to add the value "flat" for the attribute "shape" for each of our representations for "lemon", "orange", "house", "pie" and so on. Rather we would like to say that our knowledge of the "flattening" operation is generalised across our representations of categories. This notion of generalisation of knowledge of operations will be an important feature of the framework to be developed. The moral of representational economy is that we must require that whatever structures we hypothesise to describe word meanings they must genuinely generalise; in principle, they shouldn't incorporate information specific to each exemplar.

Moral 3 *The Moral of Central Exemplars*

In everyday reasoning the inferences we make regarding categories are often based on typical exemplars. If we have to explain what dogs are we tend to describe what have been called their default properties. Such an observation is made much use of by Putnam (1975) and these default properties seem to be very similar to what he calls "core facts". The pervasive nature of default reasoning suggests that the central exemplars, or those exemplars for which "core facts" are indeed facts, have special

cognitive significance. This is also something that a theory of word meaning should respect. Classical theory singularly failed to respect this moral and we might see this as the reason for its problems. Prototype theory, on the other hand, accords the central exemplars a special status in that they are represented by the prototype. Due to the weightings on attribute-value pairs, prototype theory also offers a way of distinguishing different, non-central exemplars. However, the moral of Central Exemplars is that any theory of word meaning must distinguish between the representations of central and peripheral exemplars.

Moral 4 *The Moral of Context Dependence*

In a number of ways the meanings of words can be said to depend on the context in which we find them. "Lion" when used to describe a fierce lion seems to mean something quite different from "lion" when used to refer to a statue of a lion. Typically, it appears as though the same word carries different meanings and the nature of the difference directly reflects the nature of the differences between the situations in which it is used and the purposes for which it is used. Barsalou, for example, has shown very convincingly a number of effects of context on our representations of categories (Barsalou, 1982; Barsalou, 1987). Barsalou's conclusion seems to imply that there are no stable representations corresponding to concepts (a conclusion that the framework we develop later will endorse, albeit with a qualification). However, were the same conclusion to apply to word meanings, it would appear very unintuitive: words do seem to have a stable core meaning and the meaning of words in use seems related to the core meaning. The two meanings of "lion" in the lion puzzle, for example, seem highly related. Therefore the conclusion I will take from observations of context dependence is that a theory of word meaning must have an adequate account of the role of situational factors in determining the deviation of the meaning of words in use from what I take to be their core meaning. This is the Moral of Context Dependence.

These morals allow one a considerable degree of freedom and the framework I shall present here should by no means be seen as the only approach consistent with them. As far as an account of word meaning is concerned, these are the morals we must respect:

1. Senses and concepts must express generalisations and lend themselves to economical representations *a la* classical theory. In addition, we would expect that facts about operations (such as flattening, drying, ripening, etc.) should be similarly expressed in an economical way.

2. Multiple word meanings, and hence senses, (perhaps especially those corresponding to operations) should be implicated in the various uses of a single word.

3. Our account of senses must be sensitive to the important cognitive distinction between central and peripheral exemplars. That is, we expect there to be some distinction between the senses a word has when used to describe central as opposed to peripheral exemplars.

4. The deviation of peripheral word meaning and senses from core meaning and senses should be sensitive to the situation in which the words are used and to the informational requirements of the agents involved in their use.

In describing the framework I shall offer an interpretation of it in situation theory. Though I shall not refer to it directly, an implementation in C-prolog has been completed.

4.2 Worms

We take it that situation type of uses of a given word is related to a set of properties given by that word's LEXON. We will label such a relation, WORd Meaning (or WORM). Taking the example of lemons once more, the WORM will relate the type of uses of *lemon* to a set of properties, perhaps those given in (1) which is repeated below. Our account of the Relational View will not depend on the particular properties we have chosen, that is, the particular attribute-value pairs in (1). That is, it does not depend on whether (1) does indeed describe the core sense of *lemon*. Rather, what the Relational View amounts to is a more general claim about the nature of such core senses, their relation to peripheral senses and how language users may be flexible in their use of words.

The LEXON for *lemon*, then, expresses attunement to a set of properties and their relation with certain uses of *lemon*. Let us suppose that these are the properties described below.

$$\begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{oval} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix} \tag{4.3}$$

In Situation Theory we can view this WORM as a relation in the following infon where the property P is that provided by the LEXON for *lemon*.

$$\ll \text{WORM, LEMON, P; 1} \gg \tag{4.4}$$

where

$$\begin{aligned} \text{P} = \quad [\dot{\text{p}} \mid \; & \ll \text{oval}, \dot{\text{p}}; 1 \gg \wedge \\ & \ll \text{acidic}, \dot{\text{p}}; 1 \gg \wedge \\ & \ll \text{yellow}, \dot{\text{p}}; 1 \gg \wedge \\ & \ll \text{fruit}, \dot{\text{p}}; 1 \gg] \end{aligned} \tag{4.5}$$

As suggested earlier, WORMs can be thought of in terms of constraints. One constraint might be used in parsing so as to constrain the possible interpretations associated with a given use of a word. For example, given a use of *lemon* we may want to conclude that we have an individual which has acidic taste, yellow colour and so on. Another constraint might be used in generation so as to constrain the possible ways of describing a given individual. For example, given an individual which has the properties of acidic taste, yellow colour and so on, we would wish to conclude that the word lemon may be used to describe such an individual. That is, there are two possible constraints, one which necessarily seems to involve word use, the other which involves the abstract properties of a word. Here I shall be concerned solely with the latter kind of constraint.

Our WORM will associate the type of uses of a word with a particular set of properties

provided by the LEXON for that word and so the WORM expresses the sort of constraint we need to characterise the conditions under which the word can be used. Taking the example of lemons again and assuming that we can also associate some properties with an lemon, then provided that the properties associated with the individual include those specified by the LEXON then the individual can be described as a lemon. In Situation Theory, we will represent the properties associated with the individual by means of a restricted parameter. What we then have amounts to the following: provided the restricted parameter associated with the object has those properties specified by the LEXON then that word can be used in describing the object. Following Barwise (1985), the constraint can be expressed as follows.

$$S_1 \Rightarrow S_2 \mid B \tag{4.6}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P, \dot{p}; 1 \gg] \tag{4.7}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \textit{lemon}; 1 \gg] \tag{4.8}$$

and

$$B = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LEMON}, P; 1 \gg] \tag{4.9}$$

and

$$\begin{aligned} P = [\dot{p}_1 \mid\ & \ll \text{oval}, \dot{p}_1; 1 \gg \wedge \\ & \ll \text{acidic}, \dot{p}_1; 1 \gg \wedge \\ & \ll \text{yellow}, \dot{p}_1; 1 \gg \wedge \\ & \ll \text{fruit}, \dot{p}_1; 1 \gg] \end{aligned} \tag{4.10}$$

(4.6) is a conditional constraint. That is, it is a constraint that is relativised with respect to some background situation type, B, such that provided we have a background situation of this type then the constraint holds. (4.6) and (4.9) express the fact that if we have an individual with the property P then it is in virtue of the WORM *lemon*, a component of the background conditions, that the individual can be described as a lemon[1]. Note that we have introduced another relation DESCRIBE. It is assumed that this relation holds between a use of a word such as *lemon* and an object represented by a parameter anchored to that object such that, if the relation holds, the word, *lemon*, can be used to describe the object.

So far we have a framework that is very similar to classical theory and we might consider whether it respects the morals that are so important. Although we cannot yet see how all the morals are to be respected, we can note in passing that the moral of representational economy is respected since the properties given by LEXONs express a significant degree of generalisation, *à la* classical theory. WORMs also offer a way of describing the central exemplars. The question then remains as to how to accommodate the peripheral exemplars and, as we shall see, these are to be accommodated by what we will call COMBINATIONS of WORMs (or COWORMs).

4.3 Coworms

Consider a peripheral exemplar of the category of lemons, one that has been squashed flat. That is, an individual that will not satisfy the description we gave previously for *lemon* but which would satisfy, among others, the following description.

[1] Although I shall not discuss the issue of perspectives, by locating WORMs in the background situation type this account seems to have certain interesting parallels with the notion of perspectives developed by Jerry Seligman (1990).

$$\begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{flat} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix} \tag{4.11}$$

Now, with this individual we can associate what in Situation Theory can be modelled by a restricted parameter. The parameter this time though will not have the properties given by the LEXON for *lemon* but will, instead, be as follows.

$$\begin{aligned} \dot{p} \mid\ & \ll \text{flat}, \dot{p}; 1 \gg \wedge \\ & \ll \text{acidic}, \dot{p}; 1 \gg \wedge \\ & \ll \text{fruit}, \dot{p}; 1 \gg \wedge \\ & \ll \text{yellow}, \dot{p}; 1 \gg \end{aligned} \tag{4.12}$$

Since any object to which $\dot{p}$ might be anchored will not have the properties expressed by the LEXON for *lemon*, the constraint (4.6) cannot licence uses of the word *lemon* to apply to such objects However, cognitive agents have a wealth of information at their disposal that we might characterise as being in their background resource situation. This situation allows the agent access to information about the world that will resolve certain local inconsistencies. One such inconsistency is where the object we are considering does not have the properties expressed by the LEXON for lemon (e.g., a squashed lemon) and yet may nonetheless be described as a lemon. In this instance the resource situation allows us to access WORMs that can resolve the inconsistency. One such WORM relates the type of uses of *flatten* with a LEXON giving two related properties. The properties are identical except for the stipulation that one holds of flat individuals and the other holds of individuals which are identical except that they may have any shape.

The arguments we considered in discussing coherence and representational economy indicate the kind of solution we require. We need a further constraint which expresses the fact that an individual with a certain set of properties may be described by a use of, say, *lemon*, provided that the properties it has are related, by a WORM, to those of the LEXON *lemon*. So we will have a WORM *lemon* and a WORM *flatten* which when

combined suitably will describe a flattened lemon. In Situation Theory, this can be expressed as another constraint governing the properties of the word *lemon*. Provided the restricted parameter associated with a given object is of a sort that is related by the WORM for flattened to the sort of parameter associated with the WORM for lemon, then the object is such that it could be described by a use of *lemon*. This new constraint is as follows.

$$S_1 \Rightarrow S_2 \mid B' \tag{4.13}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \tag{4.14}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \textit{lemon}; 1 \gg] \tag{4.15}$$

and

$$\begin{aligned} B' = [\dot{b} \mid \dot{b} \models\ & \ll \text{WORM}, \text{LEMON}, P_1; 1 \gg \wedge \\ & \ll \text{WORM}, \text{FLATTEN}, P_1, P_2; 1 \gg] \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} P_2 = [\dot{p} \mid\ & \ll \text{flat}, \dot{p}; 1 \gg \wedge \\ & \ll \text{acidic}, \dot{p}; 1 \gg \wedge \\ & \ll \text{yellow}, \dot{p}; 1 \gg \wedge \\ & \ll \text{fruit}, \dot{p}; 1 \gg] \end{aligned} \tag{4.17}$$

Very briefly, we have seen how this Relational View can cope with both central and peripheral exemplars. Respecting the moral of central exemplars, it treats central and peripheral exemplars differently: central exemplars are accommodated by WORMs, peripheral exemplars by COWORMs. The fact that peripheral exemplars are treated by COWORMs, however, allows the moral of coherence to be respected. For instance, in the treatment we gave for our flattened lemon, one can see the WORM *flatten* as expressing, in some sense, a theory. It is precisely this kind of "knowledge-rich" solution that the moral of coherence demanded. The moral of context dependence is also respected by ensuring that the WORMs that underly word use are seen as components of the background situation. The background situation may change either as agents focus situations change or as agents require different information from their focus situation. In addition, COWORMs capture the added generalisation that prototype theory seems to miss. Our knowledge of flattening is not encoded in each of our representations for various categories but it is encoded separately in a way that will generalise. This plausible and desirable result suggests an interesting developmental prediction. Namely, when we acquire the LEXON *flattening* we concomitantly learn that all manner of objects that are flat can be described in new ways: we learn that flat objects that otherwise look like lemons can also be described as lemons. Unlike most other accounts, this allows that we don't have to change our data structure describing lemons on encountering a flat lemon for the very first time.

In addition to respecting the Morals, I believe the account offered here is promising for other reasons. Consider a situation in which Fred comes into my office and tells me that he has a lemon in his pocket. Corresponding to an instance of default reasoning, there is a sense in which I come to believe that Fred has a yellow, oval-shaped fruit on his person. This inference can be expressed by considering that it is the LEXON for lemon which is used in such reasoning. We may, however, acquire further information that conflicts with the LEXON for lemon and so rules out *lemon* as an appropriate word. In such circumstances we may have to find a different property, which may be given by a relating this LEXON and certain other LEXON(s) via a COWORM. However, there is also a sense in which I know nothing of what Fred has in his pocket other than that there is something and Fred calls that something a lemon. The fact that so little information can

be conveyed by the use of a word seems to correspond exactly to the indeterminacy in finding a particular corresponding WORM or COWORM to describe that word's meaning.

COWORMs are also implicated in the realisation of the background conditions with respect to which conditional constraints are relativised. When we wished to describe a squashed lemon as a lemon, the first constraint (4.6) clearly failed as (4.11) does not have the property P in (4.9). The inapplicability of the constraint we can assume to be due to the fact that the actual background situation was not of type B in (4.6). The inappropriateness of the constraint leads to the utilisation of a different but related constraint as in (4.12), relativised with respect to some new set of background conditions, B′. This time the constraint holds because (4.11) has the property P_2 in (4.15). If we consider the way this constraint is expressed then we can see it as revealing something of the nature of the original background conditions with respect to which the constraint in (4.6) is relativised. The particular COWORM that is part of the second constraint tells us that the negation of that COWORM was a condition of the original background situation type. That is, returning to our example, the fact that the WORM "flattened" relates the constraint (4.12) to the constraint (4.6) tells us that (4.6) presupposes that lemons aren't squashed flat and this then is one of the background conditions with respect to which (4.6) was relativised. In this framework then, the process of finding an appropriate COWORM is the process of realising background conditions.

So now we have a very basic framework with which to explore word meaning. It seems to respect the morals that I have tried to argue are so important and by the use of COWORMs the account seems to offer a solution some of the problems we discussed in Chapter 1. Now, we turn our attention to some of the other problems of Chapter 1.

4.4 The Puzzles Reconsidered

4.4.1 The Lion Puzzle

Fred has been sitting on a park bench experiencing a number of situations. In one situation he claims to have seen a lion and in this we would probably agree. In another, he claims not to have seen a lion and again we wouldn't argue. Now we can suppose that Fred knows that something like the following constraint holds concerning what kind of object may be described as a lion.

$$S_1 \Rightarrow S_2 \mid B \tag{4.18}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models\!\ll P, \dot{p}; 1 \gg] \tag{4.19}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models\!\ll \textsc{describe}, \dot{p}, \textit{lion}; 1 \gg] \tag{4.20}$$

and

$$B = [\dot{b} \mid \dot{b} \models\!\ll \textsc{worm}, \textsc{lion}, \dot{P}; 1 \gg] \tag{4.21}$$

and[2]

[2]Clearly, P only expresses *some* of the information relevant to lionhood.

$$\begin{aligned} P = [\dot{p} \mid & \ll \text{animate}, \dot{p}; 1 \gg \wedge \\ & \ll \text{four-legged}, \dot{p}; 1 \gg \wedge \\ & \ll \text{one-tailed}, \dot{p}; 1 \gg] \end{aligned} \tag{4.22}$$

That is, we may describe the default sense Fred associates with the word *lion* as follows.

$$\begin{bmatrix} \text{ANIMATE:} & + \\ \text{LEGS:} & 4 \\ \text{TAIL:} & 1 \end{bmatrix} \tag{4.23}$$

When Fred responds to the zoo-keeper, his use is in accordance with this constraint. That is, Fred *hasn't* come across an object whose description is of the sort expressed by the LEXON for *lion*, i.e., an animate object with 4 legs and a tail. So provided that the constraint Fred is exploiting is expressed by the LEXON *lion*, he has replied correctly to the zoo-keeper. But what of the art student? Fred has access to a background resource situation which supports many infons. One expresses the fact that there is a LEXON *statue*. This LEXON relates two properties to the word *statue*. One property holds of, say, animate things (the class of things modelled by statues), the other holds of inanimate things that stand in certain relations to the animate things (the class of statues). Though this doesn't accurately or fully describe the relationship between statues and the things they model, it will suffice for our purposes. In a similar fashion to before, we may combine the WORM for *lion* with the WORM for *statue* allowing the following constraint.

$$S_1 \Rightarrow S_2 \mid B' \tag{4.24}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \tag{4.25}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \textit{lion}; 1 \gg] \tag{4.26}$$

and

$$B' = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LION}, P_1; 1 \gg \wedge \ll \text{WORM}, \text{STATUE}, P_1, P_2; 1 \gg] \tag{4.27}$$

and

$$P_2 = [\dot{p} \mid \ll \text{animate}, \dot{p}; 0 \gg \wedge \ll \text{four-legged}, \dot{p}; 1 \gg \wedge \ll \text{one-tailed}, \dot{p}; 1 \gg] \tag{4.28}$$

That is, this constraint involves a different sense being associated with the word *lion*, one which we can describe as follows.

$$\begin{bmatrix} \text{ANIMATE:} & - \\ \text{LEGS:} & 4 \\ \text{TAIL:} & 1 \end{bmatrix} \tag{4.29}$$

Now Fred, whilst walking to his park bench, *has* come across an individual whose description is of this last sort, i.e., an inanimate object with 4 legs and a tail. So, provided that the constraint he is exploiting is expressed by this COWORM, Fred can rightly assert that he has seen a lion which is, fortunately, what the art student was hoping for.

What should be clear is that COWORMs are sensitive to the information states of agents. If Fred's background resource situation did not support the infon concerning the WORM for *statue* then, according to this view, he would not be able to rightly report that he

had seen a lion. The significance of this point is much more obvious when we consider Nunberg's metonymic puzzle.

4.4.2 Nunberg's Ham Sandwich

The waiter in the restaurant knows (we can assume) that *ham sandwich* is normally appropriate to describe ham sandwiches and not human beings like the customer he is serving. However, his particular background resource situation also supports the fact that the WORM relation holds between the type of uses of the word *orders*, and the properties appropriate for things of type "customer" (P_2) and for things of type "ham sandwich" (P_1). Just as before, this WORM can form a COWORM with the WORM for *ham sandwich* which allows us to consider a new background situation type as follows.

$$\begin{aligned} B' = [\dot{b} \mid \dot{b} \models\ & \ll \text{WORM}, \text{HAM\_SANDWICH}, P_1; 1 \gg \wedge \\ & \ll \text{WORM}, \text{ORDERS}, P_2, P_1; 1 \gg] \end{aligned} \tag{4.30}$$

Similar to Fred's case, we now have a COWORM which expresses a constraint relativised with respect to this background situation type. It is this new constraint that renders it legitimate for the waiter to refer to his customer as a ham sandwich. The point about this particular example that is not quite so clear in the others I have described is precisely the contextual nature of this aspect of word use. It seems reasonable to suggest that there is something about the particular background resource situation that allows the waiter to call a customer a ham sandwich. Locating such a mechanism in the background situation seems to offer a way of explaining how it would be considered infelicitous were the waiter to describe the same person in the same way in a launderette. It seems the particular COWORM we have considered would not be supported given the nature of those prevailing background conditions.

4.4.3 Lakoff's Mother and Macken's Father

The final puzzle of Chapter 1 is that of kinship terms which is discussed in the analyses of Lakoff (1987) and Macken (1990). The problem, however, is widely known: mothers, for example, are conventionally considered to be genetically related to their offspring. Adoptive mothers, though, are naturally described as mothers even though they and their adopted children are not related genetically. Similarly, as Macken's example demonstrates, fathers may be live-in but are fathers nonetheless. The Relational View I have outlined suggests an alternative analysis to those offered by both Lakoff and Macken which we will now pursue. Under this view, we might like to say that the following constraint holds concerning the sort of objects that can be described by a use of the word *mother*.

$$S_1 \Rightarrow S_2 \mid B \tag{4.31}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models\!\ll P, \dot{p}; 1 \gg] \tag{4.32}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models\!\ll \textsc{describe}, \dot{p}, \textit{mother}; 1 \gg] \tag{4.33}$$

and

$$B = [\dot{b} \mid \dot{b} \models\!\ll \textsc{worm}, \textsc{mother}, P; 1 \gg] \tag{4.34}$$

and

$$\begin{aligned} P = [\dot{p} \mid & \ll \text{female}, \dot{p}; 1 \gg \wedge \\ & \ll \text{child\_of}, \dot{p}, \dot{q}; 1 \gg \wedge \\ & \ll \text{genetics\_of}, \dot{p}, \dot{g}; 1 \gg \wedge \\ & \ll \text{genetics\_of}, \dot{q}, \dot{g}; 1 \gg] \end{aligned} \tag{4.35}$$

That is, the default sense associated with uses of *mother* can be described as follows.

$$\begin{bmatrix} \text{SEX:} & \textit{female} \\ \text{CHILD:} & \begin{bmatrix} \text{GENETICS:} & g \end{bmatrix} \\ \text{GENETICS:} & g \end{bmatrix} \tag{4.36}$$

It is this constraint that tells us that a female who is genetically related to her child, may be described by a use of the word *mother*[3]. This, however, does not help us describe adoptive mothers: we have no mechanism for stating that an individual described by the following restricted parameter can be described by a use of the word *mother*.[4]

$$\begin{aligned} \dot{p} \mid & \ll \text{female}, \dot{p}; 1 \gg \wedge \\ & \ll \text{child\_of}, \dot{p}, \dot{q}; 1 \gg \wedge \\ & \ll \text{genetics\_of}, \dot{p}, g_1; 1 \gg \wedge \\ & \ll \text{genetics\_of}, \dot{q}, g_2; 1 \gg \end{aligned} \tag{4.37}$$

In a similar fashion to before, we will use some other LEXON to form a COWORM which provides a further constraint on the way in which the word *mother* can be used. We will assume that agents have access to the LEXON for *adoptive*. The WORM for *adoptive* relates the type of uses of the word *adoptive* and two related properties. One property holds of two individuals having, among other properties, the same genetic material, the other property is identical with respect to these other properties, except that it holds of individuals whose genetic material is unrelated[5]. This WORM, combined with the WORM *mother* gives us a COWORM that expresses the following constraint.

[3] As described here, we have said that the child and mother must have the *same* genetics. Though this is most certainly not the correct condition on the use of the word *mother* it will suffice for illustrative purposes.

[4] In (4.37), g_1 and g_2 are assumed to be distinct.

[5] Again, this is not the correct condition but will suffice for illustrative purposes.

$$S_1 \Rightarrow S_2 \mid B' \tag{4.38}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \tag{4.39}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \textit{mother}; 1 \gg] \tag{4.40}$$

and

$$\begin{aligned} B' = [\dot{b} \mid \dot{b} \models\ & \ll \text{WORM}, \text{MOTHER}, P_1; 1 \gg \wedge \\ & \ll \text{WORM}, \text{ADOPTIVE}, P_1, P_2; 1 \gg] \end{aligned} \tag{4.41}$$

and

$$\begin{aligned} P_2 = [\dot{p} \mid\ & \ll \text{female}, \dot{p}; 1 \gg \wedge \\ & \ll \text{child\_of}, \dot{p}, \dot{q}; 1 \gg \wedge \\ & \ll \text{genetics\_of}, \dot{p}, \dot{h}; 1 \gg \wedge \\ & \ll \text{genetics\_of}, \dot{q}, \dot{g}; 1 \gg] \end{aligned} \tag{4.42}$$

Now, this constraint suggests that a different sense of *mother* may be associated with its uses, namely the one described as follows.

$$\left[\begin{array}{ll} \text{SEX:} & \textit{female} \\ \text{CHILD:} & \left[\ \text{GENETICS:} \quad g_1 \ \right] \\ \text{GENETICS:} & g_2 \end{array}\right] \tag{4.43}$$

It is this constraint, expressed by the COWORM of our WORMs for *adoptive* and *mother*, that allows that individuals who are adoptive mothers and have no genetic relation to their children can nonetheless still be described by a use of the word *mother*. Importantly, this analysis does not lead to the suggestion that there are several independent models underlying the word *mother* as Lakoff's analysis does. Nor does it require that we have to accept the apparently unintuitive suggestion of Macken's analysis, that the "father" relation is really a relation between a father, a child and some parameter.

4.5 Characterising Analyticity

So far in this Chapter, we have argued that the relational view of word meaning is consistent with a number of concerns which figure in both the psychological and philosophical literatures. However, we have said little about how this view of word meaning or, more precisely, this view of senses, may be seen in terms of a wider theory of semantics. In this section, we focus on one aspect of this issue, namely the characterisation of meaning on which the relational view rests. We expand on this by initiating a discussion of analyticity. In Chapters 5 and 6, we take up this issue once more. Our attempt at addressing this important issue will focus on the relation between this theory of sense and traditional theories of meaning and, in particular, the implications of the former for the latter.

Our accounts of Sense Generation and of the Relational View have made a number of assumptions. Firstly, we have assumed that the conventional uses of words betray multiple meaning relations and, consequently, multiple senses. Secondly, we have suggested that these senses are related to Situation Theory's notion of informational content and, indeed, that these senses constitute information conveyed by utterances of their corresponding words. Thirdly, and perhaps most importantly, we have suggested that these meaning relations and these senses are descriptional. Our position, outlined in Chapter 2 and adopted in Chapter 3, is that the rejection of the general claim that word meanings may be descriptional is one that arises as a corollary of a particular approach to meaning. We suggested that it arises from construing the uses of words as betraying

one single sense, one single meaning relation and, further, that this meaning relations and, hence, the sense exhibited the property of monotonicity or indefeasibility. We argued that this construal, together with a consideration of counterfactual conditionals, prohibits meaning relations from being descriptional in general. In particular, the suggestion has been from Kripke and Putnam that natural kind terms are non-descriptional in this sense: their senses cannot be explicated in terms of descriptions.

But it should be clear that this position is in stark contrast with the one we have developed in this Chapter. With a natural kind word such as *lemon* we have associated descriptions such as *yellow, acidic, oval-shaped fruit.* The issue to which we must now turn, then, is the motivation for the position that natural kind terms are non-descriptional. For, if we are unable to combat this position, our own will surely crumble.

Now the claim that words are non-descriptional does not arise solely in the case of natural kind terms; it also arises in the case of proper names. Though proper names have been deemed to be descriptional, due largely to an influential rejection of this position by Kripke, the view that their meaning cannot be explicated in terms of descriptions is popular. Indeed, the literature on proper names is germane for our purposes for this very same rejection is the one that has ben advanced as a rejection of the view that natural kind terms are non-descriptional. The rejection is made on the basis of a traditional conception of analyticity.

Analyticity is fundamentally a semantical notion. A statement which expresses an analytic truth one that is deemed to be true in virtue solely of its meaning, that is, true almost by definition. Accordingly, assuming that a given word has a meaning, it should be the case that there is a statement which exemplifies that word's meaning such that the statement expresses an analytic truth. For example, consider our word *noj*, and suppose *noj* means *tap*, then the statement *noj is tap* will be true solely in virtue of it meaning. That is, it will express an analytic truth. Over the next two chapters we will focus on some conceptions of analyticity and, in particular, how one conception of analyticity has been used to reject the general position that the meanings of words may be descriptional. We will suggest an alternative conception of analyticity based on the

construal of meaning relations that we have adopted. That is, our construal in terms of conditional constraints.

4.6 Conclusion

During the course of this chapter we have established a particular view of senses. Specifically, we have suggested that all uses of a word may, in principle, belie a number of different senses. That the nature of these senses depends upon the context in which uses are made and that their number may, in principle, be non-denumerable. What distinguishes this position from *Sense Creation* is our treatment of conventional uses of words. In particular, we have suggested that it is logically possible to describe the meanings associated with the conventional uses of natural kind terms. That is, we have suggested that such terms may be descriptional.

Now, it is generally assumed that the arguments of Kripke (1972) and Putnam (1975) have been influential in establishing the position that the meanings of terms such as natural kind terms are non-descriptional. That is, the meanings of such terms, inasmuch as they can be said to have meanings, cannot be given by description, though this is not to say that descriptions play no role in a wider theory of meaning. However, as is clear, this is quite at odds with the relational view. For it is a conclusion of this view that such terms have multiple senses, these being amenable to description. The consequences of such a position we no more than pointed to earlier, however this striking divergence will form the focus of our attention for the next two chapters. First, we will consider the literature on proper names where the argument has also been made, contrary to established views, that proper names are similarly non-descriptional. We trace the line of argument back to a characterisation of analytic statements, statements whose truth is deemed to be both necessary and *a priori* knowable. It is this view which is at odds with the relational view we have established and an important aspect of this chapter will be to make two suggestions: that the Kripkean argument concerning proper names can be treated as a *reductio*; and that the view of analyticity we sketched in the previous section is, at least *a priori*, a plausible one.

In chapter 6 we turn to the literature on natural kind terms and show how the traditional conception of analyticity has been used to argue that these are also non-descriptional. Our argument with respect to this position, however, is similar to the position adopted on proper names, namely that such a conclusion is a *reduction* on the traditional conception of analyticity. Finally, we show how our revised conception of analyticity offers a potential solution to the problems raised by natural kinds and argue that this allows a reconciliation between psychological and philosophical views which the traditional conception of analyticity had previously prevented.

Chapter 5

Proper Names and Analyticity

> But if someone wished to say: "There is something common to all these constructions — namely the disjunction of all their common properties" – I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread — namely the continuous overlapping of those fibres".
>
> *(Wittgenstein, PI, §67)*

Having described Sense Generation, in this chapter we examine some of the philosophical literature on proper names and by this we mean, ordinary proper names. We do not mean Fregean proper names, Russellian singular terms or the like. Simply, good old-fashioned proper names. Our overall objective is to examine views of meaning which may have points of contact with our view of the generation of senses. Though this may seem something of a departure, our goal is to effect a better appreciation of Sense Generation in light of the traditional arguments which have been advanced with respect to word meaning. So we must turn to such arguments and where better to start than by examining proper names. Again, our intention is not to provide an exhaustive precis of the literature, but rather to consider trends therein.

One such trend is pointed to by Kripke's arguments and rightly so, since his position is one of refuting the efficacy of this trend. Our position is somewhat similar but our presentation will not be nearly so bold. We will argue that the Kripkean argument may be taken either of two ways. Either as an argument to the effect that the meaning

of proper names, inasmuch as they can be said to have meaning, cannot be given by description. Or, and this is what we shall suggest, as suggesting the possibility of viewing Kripke's argument as an inadvertent *reductio* on the conception of analyticity which it assumes. And further that a view of the analytic based on the picture of Sense Generation offers the possibility of rescuing description theories of proper names. With this position established, we turn in Chapter 6 to the psychological literature on concepts, focussing on natural kinds, and argue that a similar rescuing can be performed.

We begin by outlining those theories which Kripke assimilates to the category of description theories of proper names. Thus, we look first at Frege, then at Russell and then we turn to the theories of Searle and Strawson. Though, in this chapter, we will do little more than point to the fact, these description theories have striking parallels with psychological conceptions of meaning. Indeed, we will be at pains to point out that description theories often take as motivation, explanations of the epistemological status of proper names and, in this, they are quite consonant with psychology.

However, Kripke finds fault with these theories. Indeed, the problem is that that they presume the meaning of proper names can be explicated in terms of descriptions which hold only contingently, if at all. To see why this is problematic, we turn to Kripke's exposition in which the importance of necessity in naming is argued and we briefly sketch Kripke's position on proper names. Having done so we will suggest that this view leaves little room for an explanation of the epistemology of proper names, nor of the psychology of their use. However unsatisfactory this conclusion may appear, the logic of Kripke's argument seems difficult to fault.

However, as we have already mentioned, another way of viewing the Kripkean argument is as a *reductio* on the notion of analyticity on which it is based. To advance this position, we will emphasise that a theory of proper names must explain the cognitive significance of their use. Since Kripke's approach offers little hope of such an explanation, we must consider his conclusion absurd. In offering an alternative conception of analyticity, we will consider some conceptions which have already been mooted. We suggest that Kripke's arguments in favour of distinguishing metaphysical, epistemolog-

ical and semantic notions of truth, counts against the conception of analyticity which he assumes. Then, we spell out the notion of meaning which is inherent in Sense Generation, and, on this basis, suggest an alternative conception of analyticity. Finally, we will attempt to relate the picture of meaning in Sense Generation to the later positions of Frege and of Wittgenstein.

5.1 Description Theories

Kripke's seminal article, *Naming and Necessity*, offers many insights into what we should expect of a theory of meaning. One insight surfaces in his emphasising the importance of distinguishing two claims often associated with theories of meaning. One claim is that the theory is truly a theory of *meaning*, so talk of synonymy, antonymy and other meaning relations is germane to the analysis of such a theory. Another claim is that the theory of meaning is a theory of the *determination of reference*. That is, such a theory may explicate how it is that we determine the reference of various terms in our language. We might add a further distinction: Kripke talks of *fixing a reference* and, at least implicitly, he distinguishes this from *determining* a reference. At least in ordinary parlance, the fixing of reference can be taken to refer solely to the initial "long-ago" act of naming. The determination of reference, however, in its ordinary interpretation, seems to refer to all those "here-and-now" cases in which, given a name, we determine which individual is its referent.

Many of the other insights Kripke offers, we will discuss in section 5.2 when we consider Kripke's argument against description theories of proper names. However, with some of these distinctions in mind, let us try and say what is meant by a description theory of proper names, and discuss later what kind of a theory of meaning it might be.

Kripke offers (pp. 280-281) a definition of what a "cluster-concept" theory amounts to. In modified form, such a definition also serves to indicate what description theories are in general. Here, then, is this modified definition.

1. To every name or designating expression 'X', there corresponds either

   (a) a single property ϕ or

   (b) a cluster of properties, namely the family of properties ϕ

   such that the speaker A believes 'ϕX'.

2. Either

   (a) this single property ϕ or

   (b) one of the cluster of properties ϕ, or some conjointly,

   is believed by the speaker A to pick out some individual uniquely.

3. If

   (a) this single property ϕ or

   (b) most, or a weighted most, of the ϕ's

   are satisfied by one unique object, y, then y is the referent of 'X'.

4. If

   (a) the single property ϕ is not satisfied or

   (b) the "vote" between the cluster of properties ϕ yields no unique object

   then 'X' does not refer.

5. The statement,

   (a) 'If X exists, then ϕX or

   (b) 'If X exists, then X has most of the ϕ's'

   is known *a priori* by the speaker.

6. The statement,

   (a) 'If X exists, then ϕX' or

   (b) 'If X exists, then X has most of the ϕ's'

   expresses a *necessary* truth (in the ideolect of the speaker).

As we shall see, for Kripke, 6 need not be a thesis provided that the description is taken not as an explication of meaning but as a theory of the determination of reference. However, with some of the preliminaries out of the way, we can now turn to some of the theories which fit, more or less, this definition.

Before doing so, though, it is worth concentrating a little on some of the motivations for description theories of meaning. Firstly, if we explicate the meaning of a proper name by a description, then problems concerning the *meaning* of various statements such as identity statements with proper names and existential statements become soluble. Frege's analysis, for example, is often credited with being able to account for the cognitive difference between statements such as *Scott is the author of Waverley* and *Scott is Scott.* Russell's analysis is also credited with accounting for the meaning of statements such as *Pegasus does not exist.* Both these accounts depend on being able to give the meaning of a proper name in terms of some description(s), descriptions which the language user can know or come to know. The accounts of Searle and Strawson can similarly be seen as attempts to explicate, if not the meanings of proper names, their *use* in terms which make epistemological sense. Searle, for example, explicitly attempts to tie the uses of proper names such as *Aristotle* to facts about Aristotle that are known to language users. Thus, various aspects of the meaning and use of proper names are to be explicated in epistemological terms.

The determination of the referent of a proper name, then, is that process by which a language user determines which individual satisfies descriptions which the language user knows to be true of the bearer of the name. The "passing" of the referent of a proper name is then a process by which language users communicate what they know of the bearer of that name. And semantic facts concerning the meaning of proper names and the meanings of sentences in which they are embedded reduce to facts about the epistemology of language users. *Prima facie*, such a view offers considerable scope for facts about the psychology of language users to play a role in theories of the meaning and use of natural language. It is also quite consistent with our discussion in Chapter 1, where we associated with uses of words certain *conditions of use* and claimed that the satisfaction of these conditions is what makes such uses informational. The claim that

such conditions can be described lay implicit in that discussion. In the rest of this chapter, in many respects, it is precisely this that is explicit. In the next few sections we outline some commonly held description theories of proper names before turning in section 5.2 to the Kripke's argument which suggests such theories are neither theories of the meaning of proper names nor of the determination of their reference. It also suggests that the conditions of use associated with proper names may be ones that are not amenable to description. If this is correct then the possibility of explicating the uses of words in terms which make epistemological and psychological sense would appear to be remote indeed.

5.1.1 Frege

A key objective for any semantic theory is to explain facts about the significance of language in terms of facts about the semantic values that are accorded expressions in the language. As Evans (1976) tells us, in Frege's early view of language, he equated semantic value with Meaning.[1] Meaning is what we might ordinarily think of in terms of reference. So, proper names, for example, have as their semantic value a Meaning, namely, the objects which we would normally call their referents. So the Meaning of the proper name "Allegri" is the composer Allegri.

Now, as Evans points out, this view of proper names ignores much. In particular, one thing it ignores is the possibility that proper names don't correspond to objects: the name "Dracula" doesn't refer to any object, at least, no existing object. Frege, before he introduced the notion of Sense, however, was happy to concede that sentences which contained such empty singular terms were neither true nor false (Evans, p. 11). But, though this didn't particularly worry Frege, it might worry us and, indeed, Frege *was* worried by one important consequence of his theory of meaning. Namely, how to explain the cognitive significance of language.

For Frege, explaining the cognitive aspects of language was an important objective of his theory of meaning. The problem with this early theory of Meaning, however, was that

[1] We will follow Evans in using Meaning for Frege's *Bedeutung*.

it failed to distinguish between statements with the same semantic value yet different *informational* value. As Evans puts it,

> He [Frege] found it necessary to recognise the possibility of an objective semantic difference between two expressions not distinguishable by the theory of Meaning (two expressions with the same Meaning) — this difference having to do with the different ways in which the expressions are to be understood by competent speakers, these different ways in turn ultimately resting upon the different thoughts and propositional attitudes that competent speakers will have on hearing and understanding sentences containing the two expressions.
>
> *(Evans, 1976; p. 13)*

An example of the sort of expressions Frege found troublesome for his early theory is the case of two intrepid explorers who sight the same mountain from different vantage points. On their maps, they enter the name of this mountain as Mt. Aphla and Mt. Ateb, respectively but when they compare their maps, they find out that Ateb is Aphla. Now the proposition expressed by *Ateb is Aphla* is not just a consequence of the principle of identity: it does not simply assert that this mountain is self-identical. Rather it contains geographical knowledge, which has significance for the epistemological states of our two explorers. To reiterate, *Ateb is Aphla* contains knowledge over and above that deriving from the principle of identity in a way that *Ateb is Ateb* does not. Similarly, the statements *Ateb is at least 5000 metres high* and *Aphla is at least 5000 metres high*, can express different information and, indeed, the same person may take different attitudes to these statements.

Now such observations as these are problematic if one holds, as the early Frege did, to the following two principles:

1. Semantic values are to be equated with Meaning, i.e, objects of reference.
2. Semantics, built on such semantic values, must explain the cognitive significance of language.

The case of Ateb/Aphla indicates the tension between these principles. *Ateb* and *Aphla*

refer to the same object and, by 1, must have the same semantic value. But then we cannot account for the cognitive difference between statements containing these names, statements like *Ateb is at least 5000 metres high* and *Aphla is at least 5000 metres high*, for example. For such a difference in cognitive attitude must surely be because the names themselves are associated with different cognitive attitudes. Assigning them to have the same semantic value prevents this cognitive difference being drawn as a semantic distinction. And this is in direct opposition to 2. So, to maintain consistency, one or other of these principles must go. Frege, fortunately, chose to reject 1. And it is to the alternative, in favour of which he made this rejection, that we now turn.

In *On Sense and Reference*, Frege spelt out his alternative theory based on the intertwined notions of Sense and Meaning. Such problematic cases as that involving Ateb and Aphla are now to be explained by the fact that, though these proper names have the same Meaning, they have different Senses.

> It is natural, now, to think of there being connected with a sign ..., besides that to which the sign refers, which may be called the reference [Meaning] of the sign, also what I should like to call the *sense* of the sign, wherein the mode of presentation is contained.
>
> *(Frege, 1977; p. 57)*

We can obtain an idea of the intuition behind Frege's position by examining some of the other things he says in connection with senses. For example, he considers the case of a triangle where a, b and c are the lines connecting the vertices of the triangle with the mid-points of the opposite sides. Then the point of intersection of b and c is the same as the point of intersection of a and b. That is, there are different ways of designating the same point, there are different ways of presenting the same object and, in this, lies the analogy with senses. Expressions not only designate a Meaning but they present the Meaning in a certain way. This is what senses do: they present the Meaning in particular ways and this is what is meant by saying that the sense contains "the mode of presentation" of the Meaning.

So, the phrases *the evening star* and *the morning star*, though designating the same

planet, Venus, nonetheless have different senses. Just as the planet Venus presents itself to us differently according to circumstance, so the different senses of these phrases present the Meaning differently. And the same is deemed true of proper names.

Scott is Scott

Scott is the author of Waverley

It has become a common-place that the pairs such as that above betray different cognitive significances. That is, it is possible to take different attitudes to statements made with such sentences. And this despite the fact that the proper name *Scott* and the definite description *the author of Waverley* denote the very same object. Now, according to Frege, what explains the cognitive, informational difference between these two sentences is the fact that the definite description *the author of Waverley* and the proper name *Scott* have different senses. That is, they do not simply denote but they denote in a particular way, they present their Meaning in a particular way. Frege also claimed that senses are grasped by "everybody who is sufficiently familiar with the language" (Frege, 1977; p. 57). That is, they are *public*. Frege, of course, also had in mind that the sense of an entire sentence, what he called a thought, could be built, according to the principle of compositionality, out of the senses of its parts. To get a firmer grasp of what this notion of sense might mean, let us return to Evans.

Evans mentions the fact that Frege spoke of senses as "illuminating the Meaning from different sides" (Evans, 1976; p. 15). But he dismisses this and the metaphor concerning "modes of presentation" as unilluminating. Instead, he ties Frege's conception of sense tightly to his interest with thoughts.

It is unsurprising that we should turn to thoughts since, in many respects, this lies right at the heart of Frege's enterprise. By choosing to attempt an explanation of the "cognitive significance" of language, he must concern himself not only with the objects of thought, but also with the nature of the thoughts themselves. For what is revealed by cases like that of Ateb/Aphla is the fact that people may coherently take different

attitudes to one and the same sentence. In Frege's terminology, they judge the True or the False differently and to explain this fact we must concern ourselves not just with the truth value of a sentence, but how one advances to that truth value. And this process of advancing *to* a truth value involves advancing *from* a thought.

Now, as Evans notes, the notion of sense has intimate links with propositional-attitude psychology. For Frege, the sense of a sentence, composed of the senses of its parts, is a thought. And the single constraint that Frege imposed on this conception of thought is, as Evans calls it, "the Intuitive Criterion of Difference":

> ...the thought associated with one sentence S as its sense must be different from the thought associated with another sentence S' as *its* sense, if it is possible for someone to understand both sentences at a given time while coherently taking different attitudes towards them, i.e., accepting (rejecting) one while rejecting (accepting), or being agnostic about, the other.
>
> *(Evans, 1976; pp. 18–19)*

But, then, what might the sense of a proper name actually be? It is clear from the above that the sense of a proper name is the contribution which the name brings to the thought of any sentence in which it is found. For example, the thoughts associated with the sentences *Ateb is at least 5000 metres high* and *Aphla is at least 5000 metres high* are different, by the above criterion. So, correspondingly, must the senses attached to the proper names. But these different senses capture the fact that the names *Ateb* and *Aphla* while referring to the same object, are attached to different thoughts about this object. And it is this that leads Evans to claim that the different senses of proper names are really different "ways of thinking" about an object. So Frege's claim that senses must be public or shared amounts to the claim that people who understand a proper name must not only think of the referent of the proper name, but that they must all think of the referent in the same certain way. And the failure to take the same attitude to the Ateb/Aphla sentences implies a difference in the ways in which someone is thinking of the same object.

Now, Kripke viewed Frege's position as amounting to a simple description theory. And, indeed, although Evans suggests Frege was inexplicit on this point, he concurs.

> It is fairly clear that he [Frege] supposed that the way in which we think about a great many objects is, as Russell would say, "by description".
>
> *(Evans, 1976; p. 18)*

As we have already mentioned, there seem to be numerous advantages to treating proper names in terms of descriptions. One, as we have seen, is a treatment of cognitive significance in terms of semantic value, a treatment which has its roots in propositional-attitude psychology. By linking senses with descriptions, however, Frege's goal is widened. For it represents an explicit attempt to introduce into a theory of semantics, descriptions, contingent facts of which language users are aware. That is, one can see Frege's enterprise, while being overtly anti-psychologist (he claimed that senses are public, not private), is committed, if not to a wedding of semantics and epistemology, at least to an engagement.

5.1.2 Russell

Russell's conception of non-logically proper names is, quite simply, that they abbreviate definite descriptions. This leads Russell to distinguish expressions which refer by description and which need not, consequently have a referent, from those which refer directly, and must therefore have a referent. The latter are commonly called Russellian singular terms and are thought to include demonstratives such as *this* and *that*. Our interest, however, is with Russell's analysis of ordinary proper names, which he considers to refer by description.

Russell held the view that empty singular terms are only Russellian singular terms if they fail to express a proposition when in subject position. In *Principia Mathematica*, we find:

> Whenever the grammatical subject of a proposition can be supposed not to exist without rendering the proposition meaningless, it is plain that the grammatical subject is not a proper name [in the Russellian sense], i.e., not a name directly representing some object.
>
> *(Whitehead & Russell, 1927; p. 66)*

Accordingly, what are commonly thought of as empty singular terms, *Pegasus*, *Superman*, etc., cannot be thought of as directly referential terms. For, according to Russell, when in the subject position of some sentence, the failure of such terms to refer, renders the sentence an aberration: someone uttering such a sentence would be like someone uttering nonsense. So, patently, names such as *Pegasus* cannot be of this type: they cannot directly refer and so, according to Russell, they refer by description. A name such as Pegasus, for example, might refer via the satisfaction of a description such as *the winged horse of Greek mythology.* Indeed, such a description is taken to not only determine the reference of the name, but also to explicate its meaning. That is, *Pegasus does not exist* means the same as *The winged horse of Greek mythology does not exist.* Russell's analysis of proper names, then, boils down to an analysis of definite descriptions. And in this there are many advantages.

One advantage is the treatment of positive and negative existential statements. A statement such as *Pegasus does not exist* first reduces to the statement *The winged horse of Greek mythology does not exist.* This further reduces, given Russell's analysis of definite descriptions, to the statement *It is not the case that there is a winged horse of Greek mythology and there is no more than one winged horse of Greek mythology*, which is clearly true.

So, with respect to proper names, Russell's view is not unlike Frege's. Certain ordinary language proper names can be meaningful, alternatively, they can have sense, while being, nonetheless, empty. And, indeed, if we advance the thesis that senses can be explicated by description, then the picture we have is clear. Both Frege and Russell adhered to description theories of proper names, both as theories of meaning and as theories of the determination of reference. The thesis is one, however, that should present us with no difficulty, for it is one that Kripke is most explicit about.

> Frege and Russell both thought,..., that Mill was wrong in a very strong sense: really a proper name, properly used, simply was a definite description abbreviated or disguised.
>
> *(Kripke, 1972; p. 255)*

Now, according to Evans' interpretation, this is a little strong, for Frege never explicitly equated senses or ways of thinking of an object with descriptions. Nonetheless, this is the way in which Frege has been standardly interpreted. So let us just summarise some of the advantages we have observed with these description theories of proper names. Firstly, they offer an explanation of how reference is determined and, further, they do so in terms which have direct connections with the epistemology and psychology of the use of proper names. Secondly, as we demonstrated with Frege's analysis, an account of the informational differences between various identity statements is forthcoming. And finally, as we have just seen with Russell, description theories allow for an account of existential statements.

Before we consider Kripke's arguments against the Frege-Russell position, we will briefly consider some of what Kripke calls "cluster-concept" theories. The reason for the advancing of such theories is, in many respects, to combat the epistemological naivety of the Frege-Russell position.

The Frege-Russell approach assumes that associated with a proper name there is some description; a description which it is possible for a language user to know, and which, in all likelihood, the language user does know. Searle and Strawson, however, take this to task. For, according to this picture, it is simply not the case that all language users will grasp the same sense of *Aristotle*. Some people may not know of Aristotle that he taught Alexander, others will not know that he wrote De Anima.[2] In order to reconcile description theories of proper names with the facts about the epistemology of language users, the simple picture which associates a single description with a single name must be revised. The revision that Searle and Strawson suggest is that a proper name is associated with a cluster of descriptions, only some proportion of which need be known by an individual language user. Let us consider their suggestions in turn.

[2]Though it is true that Frege recognised the possibility that different language users may associate different senses with the same name this was not a property that he felt an ideal language should exhibit. Indeed, though he raises this possibility, he does so only in a footnote. Our point, in opposition to Frege, is that the fact that the same word may possess many different senses is not a sloppiness of a less than ideal language, but a central aspect of our own natural language.

5.1.3 Searle

Searle's approach to proper names is to deny the efficacy of a strict sense-reference (sense-Meaning) distinction, a distinction which the positions of Frege and Russell assume. For Searle, the whole point of names is that they avoid having to establish identifying descriptions for the objects to which they refer. It is the difficulty of establishing such descriptions that leads languages to give the whole enterprise up as a bad job, and plump for good old-fashioned proper names instead. But, we cannot exclude descriptions from playing any role whatever in an analysis of proper names. For as Searle also notes, to say that *Aristotle does not exist* is not to say that the name *Aristotle* has no denotation; rather, it is to say, that the name does not have a certain type of denotation. In some sense, then, we would want to say that a negative existential denies some descriptive statements, descriptive statements associated with the proper name.

Thus, there is a considerable tension between descriptive theories of proper names and what we might call denotative theories, those which hold that reference is direct or Russellian. Searle sees the possibility of resolving this tension by relegating descriptive content to the level of the conditions of use of a proper name rather than the level of meaning. For example, he says "unlike definite descriptions, they [proper names] do not in general *specify* any characteristics at all of the objects to which they refer". That is, a definite description refers by specifying characteristics: "the teacher of Alexander" refers to an individual by specifying that the referent of such a description has the characteristic of being Alexander's teacher. Aristotle, on the other hand, does not work that way. It specifies no characteristics of Aristotle, merely that the referent *is* Aristotle.

But, the question will always be put, how is the reference achieved? Or, in our terminology, how is it determined? According to Searle, proper names are such that "their referring uses nonetheless presuppose that the object to which they purport to refer has certain characteristics". And he goes on to say that these certain characteristics must satisfy a sufficient but, so far, unspecified number of descriptive statements which are true of that object. So, the picture of proper names that Searle paints, is the following:

associated with a proper name is a set of statements which purport to be about the referent of that name; the referent of that name is, then, that unique individual which satisfies a certain proportion of said statements. Indeed, Searle goes further: he says "...I am suggesting it is a necessary fact that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him: any individual not having at least some of these properties could not be Aristotle".

Consequently, we should best interpret Searle's account, not as a theory of the *meaning* of proper names like the positions of Frege and Russell, but as a theory of the determination of their reference. And in this respect it is very like the view of Strawson, to whom we now turn.

5.1.4 Strawson

The account of proper names that Strawson (1959) gives is so similar to that of Searle's that we will not dwell on it long. The account we relate, however, is based on Strawson's *Individuals*. Strawson suggests that associated with a proper name such as *Socrates* is a set, a presupposition-set, of propositions such that the referent of the name satisfies some proportion of this set. To understand better the nature of the proposal we need to consider the epistemological origins of Strawson's account.

According to Strawson, the successful common determination, by speaker and hearer, of the referent of a proper name depends on the speaker and hearer being able to cite facts which are true of the putative bearer of the name. Needless to say, speaker and hearer do not need to have in mind the same set of facts. Indeed, such a restriction, if it were an actual one, would very nearly prevent proper names from being used as part of the common language. For the epistemological conditions of language users can hardly be said to be common. But, nonetheless, Strawson explicates his account of proper names along the lines of their being *some* common epistemological condition of successful users of a proper name. Suppose we take a group of people who believe they use the same name to refer to the same individual. And, further suppose that

each of these people writes down a list of salient facts pertaining to the individual that they take to be the bearer of that name. Pooling these facts so as to incorporate the most frequently mentioned facts will result in a set of propositions, a presupposition set. Then Strawson's claim is that, while it may not reasonable to say that the bearer of the name is that individual satisfying *all* the propositions of that set, it is reasonable, Strawson suggests, that the bearer of the name be that individual satisfying some certain *proportion* of the propositions of that set.

Strawson also suggests that specifying the precise proportion of propositions which need to be satisfied by an individual for that individual to be the referent of the name, is, in general, not something that can be done precisely. So the theory lacks a certain amount of precision. However, what may be said in its defence is its commitment to tying epistemology to semantics. This account of proper names, which is, properly, a theory of the determination of reference of proper names rather than a theory of their meaning, explicitly integrates facts of which language users can manifest an awareness into an account of reference. As such it is consonant with the general enterprise we have attempted to articulate, that of integrating facts about the psychology of language use with semantical facts such as reference.

We now turn, however, to an argument that rejects the efficacy of explicating such notions as reference and meaning in terms of descriptions and, hence, of the tying of semantics to epistemology. That is, we turn to Kripke.

5.2 Naming and Necessity

Kripke's position on proper names is in opposition to all those theories we have so far discussed. That is, in opposition to description theories, be they theories of meaning or theories of the determination of reference. And the reason for Kripke's opposition can be found in the title of his article: the role of necessity in a theory of naming.

Firstly, Kripke convinces us of the need to be clear about a number of notions that often

appear in formulations of theories of meaning, notions concerning various categories of truth. One notion is that, in contrast to contingent truths, there are also necessary truths. Another is that, in contrast to *a posteriori truths*, there are also *a priori* truths. Then there are deemed to be analytic truths, which are to be contrasted with synthetic truths. Let us take these in turn.

The question of necessary truths is essentially a metaphysical one. The claim that a truth is necessary amounts to the claim that the world could not possibly have been otherwise. If the statement *gold has atomic number 79* expresses a necessary truth then the world could not possibly have been such as to make gold have any other atomic number. Alternatively, if we think in terms of possible worlds, then we would say that there is *no* possible world in which gold has an atomic number other than 79. So, claims about truths being necessary are claims about the nature of reality, about the essential structure of the world and how the world might have been. That is, they are metaphysical claims. In this sense, claiming that a truth is contingent is a weaker metaphysical claim. Contingent truths are ones which hold in some worlds and not in others or, alternatively, indicate that the world could have been otherwise. If I utter *I have read Naming and Necessity* my utterance expresses a truth but it is only contingent since I might never have done so. Thus, the claim which is made is weaker: it is not a claim about all worlds, merely about this one.

A related notion concerns how we come to know truths, in particular, whether truths can be known *a priori*: that is, whether a truth can be known independently of any experience. Often it has been thought that only necessary truths can be known *a priori* but for Kripke and for us this represents a confusion. Whereas necessity is a metaphysical notion, *a prioricity* is an epistemological one. It concerns our knowledge of things and how we come by that knowledge. For instance, meaning relations, such as entailments, are often thought to have this *a prioricity* to them. Knowing that p entails q and knowing that p is sufficient for knowing that q. No further experience or empirical enquiry is needed. And so the claim is that entailments are *a priori* relations. Perhaps this is a bad example because entailment relations are also necessary. However, the claim is often made in the literature on criterial relations that these are *a priori*

relations, but that they are not necessary (cf. Baker, 1974; 1977: McDowell, 1982). At any rate, it is quite clear that this notion of *a prioricity* is quite different from the metaphysical notion of necessity. Indeed, one of the important results of Kripke's article is to demonstrate this: that there are contingent *a priori* truths.

A third notion concerns analytic and synthetic truths. We will have more to say about this later (section 5.3), so for now we will consider what Kripke has to say about analyticity. Analyticity is *prima facie* a semantical notion. Statements are said to express analytic truths if those statements are true solely in virtue of their meaning. The standard example, and one Kripke gives, is that of *bachelors are unmarried men.* The truth of this statement is due not to some empirical investigation: we do not investigate all bachelors to see if each is an unmarried man. For it follows from the meanings of the words in the statement and hence of the statement itself, that the statement is true.

Now how this last notion of analyticity sits with the other categories of truth we have discussed, it seems to me, is an open question. However, for Kripke it is not: "...let's just make it a matter of stipulation that an analytic statement is in some sense true by virtue of its meaning and true in all possible worlds by virtue of its meaning. Then something which is analytically true will be both necessary and *a priori*" (Kripke, 1972; p. 264).

Thus equipped with a notion of analyticity, Kripke begins to disassemble the various forms of description theory of proper names. He begins with the Frege-Russell approach which, even though Frege's commitment to senses as descriptions was at best implicit, he regards as the view that the meaning of a proper name can be give by a definite description.

Frege gave the example of *Aristotle* having the sense of *the teacher of Alexander.* Now as a theory of meaning, which is undoubtedly as Frege meant it, such a view of the sense of *Aristotle* would have to say that *Aristotle was the teacher of Alexander* expresses an analytic truth, one whose truth is guaranteed by the meaning of the statement.

Accordingly, the issue is whether this statement also expresses a truth which is both necessary and *a priori* for this is how Kripke has construed analyticity.

Establishing whether the statement expresses an *a priori* truth is difficult. For the question must be, *a priori* for whom, since the notion of *a prioricity* is a modal notion: not that something must be known independently of experience, but that it *can* be known independently of experience. So, the fact that you or I may not have known whether Aristotle taught Alexander presumably has no bearing on the matter. It does not show that the truth of the statement cannot be known independently of experience. So, for sake of argument, let us suppose that the statement does express an *a priori* truth.

Now does it express a necessary truth? Remember that necessity is a metaphysical notion, to do with how the world might otherwise have been. So the question reduces to: could the world have been such that Aristotle was not the teacher of Alexander? Well, Kripke's answer is that this is obviously the case: Aristotle might never have met Alexander, he might never have entered into pedagogy. So Kripke's claim is that the statement cannot express a necessary truth and, in this, he is surely right. Consequently, the statement *Aristotle was the teacher of Alexander* cannot express an analytic truth and the description *the teacher of Alexander* cannot be the sense or the meaning of *Aristotle.*

The important question which Kripke has raised is: what would the sense of a proper name have to be like in order for it to be analytically true of the bearer of that name? And the only answer which appears to be forthcoming is that the sense would have to express the essential properties of the bearer of a proper name, for it is, by very definition, only these properties that are metaphysically necessary. The difficulties in specifying such essential properties, however, leads to at least three options.

1. Revise the notion of sense so as to capture necessity.

2. Deny that proper names have senses.

3. Deny that the analytic is the *necessary a priori.*

The first option is that taken by theorists such as Searle, Strawson and others and is one that Kripke has already considered. The second option is the one chosen by Kripke. The third option is the one we will explore in section 5.3. Let us take options 1 and 2 in turn.

As we have already noted, Searle and Strawson adopt what Kripke calls "cluster-concept" theories of proper names. That is, they hold that associated with a proper name such as *Aristotle* there is a set of statements which are true of Aristotle (though note, not necessarily true). The referent of *Aristotle* will then be that individual who satisfies some unspecified proportion of these statements. Indeed, Searle's view is that it is a necessary truth that Aristotle satisfies such a proportion. As we have seen, he claims that an individual not satisfying this proportion of statement commonly attributed to Aristotle could not *be* Aristotle.

Kripke's response to this position is in many respects similar to his response to the Frege-Russell position and we will summarise it as follows. Take this set of true statements commonly attributed to Aristotle. Most of these will only be contingently true of Aristotle, they will express facts such as his teaching Alexander, for example, or facts about his writings. Some, conceivably, may express necessary truths. Now consider all these statements together. In what sense is some proportion of them *necessarily* true of Aristotle? After all, the disjunction *Aristotle was the teacher of Alexander or Aristotle wrote De Anima* is no more a necessary truth than either of the terms of the disjunction. It is quite possible that Aristotle might, in another possible world, have done neither. That is, the disjunction of two contingent truths is itself a contingent truth.[3] The disjunction of a necessary truth with any other statement must, of course, yield a necessary truth. So the only sense in which some proportion of the true statements commonly attributed to Aristotle can be *necessarily* true is in the sense in which this

[3]Note here it is important that we are considering the disjunction of truths and not falsehoods. The disjunction *Aristotle wrote De Anima or it is not the case that Aristotle wrote De Anima* obviously expresses a necessary truth, albeit a tautologous one, but one term of the disjunction is false in this world. And so it cannot be a true statement commonly attributed to Aristotle

proportion is made up of individual statements that are themselves necessarily true. So the "cluster-concept" approach can only deliver the Kripkean picture of meaning for proper names in as much as they can explicate the necessary properties of the individual bearers of those names. Thus, the approach is hardly an extension of the Frege-Russell position at all, and just as that fails to capture analytic truths so the "cluster-concept" position does too.

Kripke's response to the failure to reconcile the view that proper names have senses or meanings with the view that analytic statements are ones whose truth is necessary is to jettison the whole notion that proper names have senses. Though earlier in *Naming and Necessity* he tells us that he does not attempt to present an alternative theory, he does, at least, detail some components such a theory may have.

Firstly, Kripke introduces some terminology. *Rigid designators*, designate the same individual in all possible worlds in which that individual exists. A *non-rigid* or accidental designator is one which does not designate the same individual in all worlds in which that individual exists. And a *strongly rigid designator* designates the same individual in all worlds, that is, the individual must also exist in all worlds, it must be a necessary existent. The preceding discussion of description theories will hopefully have made clear that Kripke views proper names as *rigid designators*. So, for example, *Thatcher* designates Margaret Thatcher in all possible worlds where she exists. However, *the Prime Minister of the UK in 1984* is not a rigid designator: in this world, it designates Margaret Thatcher; in another possible world it designates Michael Foot. And this, for Kripke, gives an alternative way of saying that proper names cannot abbreviate descriptions. Because descriptions are standardly non-rigid designators, while proper names are rigid. No amount of semantic alchemy will turn one into the other.

Now perhaps the most obvious retort to Kripke would be along the following lines. Granted, you have shown that there is a difference between names and descriptions and you have argued that the former cannot be reduced to the latter. But surely this is a mistake. For there are many good reasons to suppose that proper names have senses. Firstly, names are not contentless. To deny that Aristotle exists is not to simply

assert that the name has no denotation but rather that it doesn't have a certain *kind* of denotation. Secondly, if we assert that "Hesperus is Phosphorus" we do not simply state the law of identity. Yet a view which states that proper names have only a denotation would have difficulties explaining this. Finally, and most importantly, one motivation for senses is explaining how it is that people *determine* the referent of a proper name. This is what descriptions do: they determine which objects do and do not satisfy them.

Kripke does not discuss the issue of existential statements in *Naming and Necessity.* He does, however, discuss the issue of identity statements such as *Hesperus is Phosphorus.* The fact that such statements can be used to convey information he explains as follows: it is because, in general, they express necessary *a posteriori* truths. That is, the fact of Hesperus being Phosphorus is, indeed, a necessary fact and hence the statement necessarily true. However, this fact may be one of which we are unaware. That is, our epistemic states may be unable to distinguish the case where these two stars in the morning and evening are one and the same, from the case where they are in fact two distinct stars or planets or whatever. So, the fact is also an *a posteriori* one: it is incumbent upon us to investigate the situation before we may know it.

So Kripke does at least offer an explanation of why such identity statements are more than simple statements of the law of identity. Now let us turn to Kripke's incomplete picture of proper names and, in this, we will let him speak for himself.

> A rough statement of a theory might be the following: An initial baptism takes place. Here the object may be named by ostension, or the reference of the same may be fixed by a description. When the name is 'passed from link to link', the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it. If I hear the name 'Napoleon' and decide that it would be a nice name for my pet aardvark, I do not satisfy this condition.
>
> *(Kripke, 1972; p. 302)*

Kripke offers a little more on how this mysterious 'linking' is achieved.

> Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through

> various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman was a famous physicist. A certain passage of communications reaching ultimately to the man himself does reach the speaker. He then is referring to Feynman even though he can't identify him uniquely. He doesn't know what a Feynman diagram is... So he doesn't have to know these things, but, instead, a chain of communication going back to Feynman himself has been established, by virtue of his membership in a community which passed the name on from link to link...
>
> *(Kripke, 1972; pp. 298-299)*

So the view that Kripke presents of proper names is that there is an initial baptism in which an individual is named. This fixing of the reference of the name may be by ostension or by giving a description which identifies the individual in the situation of baptism. From then the name is passed, rather as the end of a piece of string can be passed, from individual to individual and, provided the string stretches all the way back to the individual baptised, each user of the name, each holder of the string, succeeds in referring to that individual with uses of the name.

Now, though the analogy is most appealing, one has every reason to expect that a theory of names should indicate in a little more detail how this name is passed from one individual to the next. It surely is *not* simply by uttering the name but, presumably, also by the giving of information concerning the bearer of that name. And it is precisely the role of this information that is of interest for description theories and it is precisely this that Kripke fails to discuss. But it is hardly surprising, such an account would necessarily be highly complex and involved. However, there are other quibbles with Kripke to consider.

Kripke's account is such as to deny the relation between knowledge about an individual and the proper name which refers to that individual. That is, it is to deny that proper names are *descriptional*, it is to deny that the meanings of proper names are associated with any descriptive content. The reason is simple: equating the analytic with the necessary *a priori* suggests that the only descriptive content associated with names

would be that corresponding to unknown essential properties. However, despite this reasoning, the conclusion seems to fly in the face of the facts.

Suppose, for example, that all Earthly language users used the names Hesperus and Phosphorus as if they did in fact refer to different heavenly bodies even though, as we know, they do not. Further, corresponding to these different names let us suppose there are different conditions of use, consistent with the idea that they refer to different bodies. In fact, let us suppose that our language users use Hesperus to refer to Venus in the evening and Phosphorus to refer to Venus in the morning. Now, consider a statement such as *Hesperus appeared red and fiery, yesterday.* The truth conditions are just the same as those of the corresponding statement *Phosphorus appeared red and fiery, yesterday.* But this is to ignore the crucial difference in the information conditions associated with the two statements: the former contains information to the effect that Venus appeared red and fiery yesterday *evening* whereas the latter contains information to the effect that Venus appeared red and fiery yesterday *morning.*

Explicating meaning in terms of necessity entails that the only inferences one can derive from the meaning of a word, or its use, are those to do with the *essential* properties of the referent. It is precisely the point of the previous illustration that words, even proper names, can be used to do much more. In particular, they can be used to convey information about incidental properties of the referent and this important aspect of the use of language is simply ignored by a view like Kripke's. So, in a very important sense, Kripke is wrong when he claims that the user of a proper name need not know anything in particular about the referent, for it is precisely what she does know that enables that user to extract and convey information over and above that to do with essence.

It seems, then, that the view Kripke espouses, ties too closely a view of semantics with a view of metaphysics, a view of how the world really must be. While we are not advocating a rejection of Kripke's ontologically realist stance, his view just seems to fail to appreciate the way words work. By explicitly expelling epistemology from semantics, as he has done with proper names, the possibility of explaining how individual language users can use language as an informational system is lost. This goal requires, almost by

definition, that the information carried is known, that it be used in making inferences, etc. So a view such as Kripke's which allows proper names only to carry information about essences must be considered unsatisfactory. We conclude this discussion of Kripke by echoing the sentiments we expressed in the first section of the first chapter. There, we drew a distinction between two questions: what are the conditions under which the word *X* could be used and what is an X. Kripke's account of proper names has, it seems, reduced the former to the latter.

5.3 Analyticity: Towards a Revision

A number of intuitions concerning analyticity have prefigured Kripke's discussion. Leibniz drew a distinction between "truths of reason" and "truths of fact" and further suggested that the former have the property of being necessary while the latter are contingent. Leibniz also made the suggestion that necessary truths of reason are ones that are true in "all possible worlds". Leibniz' distinction parallels a distinction made by Hume between "relations of ideas" and "matters of fact" and it is this distinction which is often thought to lie between analytic and synthetic truths. Kant's concern with this distinction led him to outline a number of characterisations of analytic truths. For him, statements that can be cast into subject-predicate form which are analytic statements are ones whose predicate is contained in, and is identical to, their subject. Consequently, analytic truths are ones which cannot be denied without expressing self-contradiction. In recent times great attention has been focussed on the validity of the distinction. Quine (1951), White (1950) and illustrious others have sought to argue that the distinction is without foundation for natural languages, while others have argued in opposition. In this section, I am concerned not so much with this last debate but with Kripke's characterisation of analytic truths as ones which are both necessary and *a priori* and the suggestion that this may not be the only way of conceiving the analytic. It is this we will now consider.

In Chapter 3 we explored an approach to senses in which the idea that they should express necessary and sufficient conditions was explicitly rejected. In many respects,

then, most of this chapter will have appeared a little askew: we have shown how Kripke has proceeded to dismiss description theories of proper names precisely because such descriptions do not express necessary truths. In order to clear this up and show how it is only an apparent mess, we will recap on the reasons why we were able to claim that the descriptions in Chapter 3 are senses.

Senses are those descriptions which express the informational content associated with various meaning relations. And meaning relations come in different sorts. Situation Theory, for example, has conditional constraints which hold only in certain circumstances. So, the information that these relations convey is not, in any sense, going to constitute necessary facts. If we take a constraint $p \Rightarrow q$ given r, then in r-situations, the relation between p and q is informational. In $\neg r$-situations, however, the relation between p and q is not informational. So, in no sense can p be considered a sufficient condition for q, nor q a necessary one for p. But this is not to deny that the relation can be informational and nor is it to deny that we can describe this informational relationship. Indeed, it is just such a description we call a sense. What this construal denies is that the sense of a term must be explicable in terms of necessary truths.

Indeed, the picture of meaning, and hence of analyticity, that Sense Generation offers is that meaning may be explicated in terms of contingent properties. That, within r-situations, it makes sense to think of these as "necessary and sufficient", though they are, in fact, nothing of the kind. Also, that such relations are *a priori*, in that, in r-situations, knowing p and knowing $p \Rightarrow q$ given r is grounds for knowing q. No empirical investigation is required. So the suggestion, then, from Sense Generation is that the analytic can be equated with the contingent *a priori*.

The view we offered in Chapter 3 was also that associated with a single word, even with the conventional uses thereof, there are multiple senses. Now, in an intriguing way, this can be seen to accord with some of the Fregean intuitions behind senses. Firstly, as Evans (1976) has indicated, Frege viewed the sense of a proper name, for example, as a way of thinking of the bearer of that proper name. The fact that the same individual, Ateb/Aphla for example, can have two names and also two associated senses, indicates

that, for Frege, such an object could be thought of in, at least, two different ways. Presumably, there are infinitely many ways of thinking of an object but, of course, the important point about senses is that these are ways of thinking that must be *shared* though, of course, this is not Frege's view of indexicals such as *I*.

This Fregean picture entices us, then, to see the different senses that we have suggested may be associated with words such as *lemon*, *father*, *lion*, etc., as indicating different ways of thinking about lemons, fathers and lions. But, indeed, in some way this is precisely what the puzzles of Chapter 1 do suggest. Fred's uses of *lion* do indicate that he is thinking of lions in two different ways. One way he thinks of lions is as real, animate, ferocious zoo-escapees and another way he thinks of lions is as statues. One way in which Richard, in Betsy Macken's example, is thinking of fathers is as biological fathers but, as Cassandra reveals, one can think of fathers as those male adults who live with you, feed you, clothe you and buy you bicycles. Of course, the Fregean picture is enticing, but we will not lay any more store by this than this little discussion.

Rather, we should now point to the logical possibility of an alternative account of proper names, one based on Sense Generation. Let us consider for our example the proper name *Aristotle*, and let us suppose that, by default, this name has the sense "the teacher of Alexander". Now, the argument that this cannot be the sense of *Aristotle* because it does not express a necessary truth concerning Aristotle need not alarm us. For our position is that senses may express contingent truths and still *be* senses. So much of the argument that Kripke presents against description theories of proper names will not affect our analysis. What may affect it, however, relates to Kripke's consideration of counterfactual conditionals. Suppose we consider the counterfactual conditional *If he had not taught Alexander, Aristotle would have become a hermit.* However implausible this conditional may be is not at issue. What is at issue is whether, in these counterfactual circumstances, *Aristotle* still refers to Aristotle.

Clearly, if the only means by which the name may refer is via the description, "the teacher of Alexander", then the name will not now refer to Aristotle. So, even allowing for the possibility that senses may express contingent *a priori* truths will not furnish an

explanation of *all* the uses of *Aristotle*. However, in Sense Generation, we are committed to the view that not just one description underlies the uses of a term. Instead, we may, according to circumstance, generate new descriptions which appropriately determine reference in those circumstances. So, in the case of *Aristotle*, we can imagine that, in the counterfactual circumstances we have described, a new description or sense of *Aristotle* is generated. Perhaps, the description "the clever hermit philosopher of ancient Greece". And similarly for other counterfactual circumstances we can imagine that, in these circumstances, a different sense from normal is generated. Even in cases where we cannot begin to specify what this new sense may be, we may, nonetheless, know enough to realise that the conventional sense is *not* appropriate and that some other, unspecified sense is appropriate.

Similarly in the case of *Hesperus* and *Phosphorus*. We can envisage these having the senses "the evening star" and "the morning star" respectively. A claim that these senses do not express necessary truths concerning the planet Hesperus/Phosphorus cannot now count against a claim that these are, nonetheless, senses. And, similarly, by allowing that, in principle, these names may, in each counterfactual situation, have a different though related sense, we allow, in principle, for an account of the counterfactual, possible uses of these names.

Now, this said, it seems that the intuition that proper names do not have senses, unlike other terms such as common nouns, is very strong. All we have attempted to demonstrate is the *logical possibility* that we may treat proper names as having senses. Indeed, by allowing for this possibility we also allow this intuition to play a more significant role in a theory of proper names. Whereas, in Kripke's account, regardless of our intuitions, we could not allow for the possibility of proper names having senses. How this intuition regarding proper names having senses is to be treated is, according to our account, open to the theorist to decide. Our account, while not claiming that proper names have sense, is agnostic as to the possibility.

We part by noting that the picture adopted in Chapter 3 has the potential to offer a better interpretation of Wittgenstein than that offered by Kripke. Kripke attributes

to Wittgenstein the belief in a "cluster-concept" position, namely, the belief that the meaning of a proper name can be given by a (single) *family*, or cluster, of descriptions. The picture offered by Sense Generation, however, is not a single cluster of descriptions but a whole host of related descriptions. That is, a *family of constraints*, each constraint being appropriate for a particular type of circumstance. We will have more to say about Wittgenstein's notion of family resemblance in the next chapter so, for now, we will simply repeat the quote with which this chapter began and let Wittgenstein speak for himself.

> But if someone wished to say: "There is something common to all these constructions — namely the disjunction of all their common properties" – I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread — namely the continuous overlapping of those fibres".
>
> *(Wittgenstein, PI, §67)*

Chapter 6

Natural Kinds and Sense Generation

In this chapter we discuss some of the psychological and philosophical literature on natural kinds. One aspect which we will consider is the way in which these literatures, though having undergone similar developments, have recently diverged. In the psychological literature, approaches to natural kind terms are found couched in one or other of the corresponding theories of natural kind concepts. So we begin by considering the so-called classical theory of concepts and relating it to the Frege-Russell conception of the meaning of proper names. The literature on proper names will prove germane for our purposes since we will contend that the same considerations that are raised to counter description theories of proper names are also ones which form the basis of the Kripke-Putnam rejection of description theories of natural kind terms. With regard to the classical theory of concepts, our position is that this fails to counter the same objection which Kripke raised against the Frege-Russell view of proper names. Namely, that natural kind terms, just as proper names, must designate rigidly. That is, their designation can only be expressed in terms of statements that are necessarily true. Since descriptions are typically contingent, the meaning of natural kind terms, so the argument goes, cannot therefore be explicated in terms of such descriptions.

We then turn to a consideration of prototype theory, the other major psychological theory of concepts. We will again relate this psychological theory to a philosophical ap-

proach to the meaning of proper names, that defined by the so-called "cluster-concept". Our position is, once again, that the main argument to count against prototype theory is that which Kripke also raises against the cluster-concept approaches. Though there are differences between the types of description associated with proper names and natural kind terms, it seems that even a cluster of contingent descriptions will not explicate what Kripke would take to be the meaning of a natural kind term.

A third approach taken within the psychological literature is to adopt a so-called hybrid approach to concepts in which conceptual cores are to be distinguished from identification procedures, the former being accommodated by something akin to classical theory, the latter by something akin to prototype theory. Our position with respect to these theories of concepts is that they fail to overcome the deficits of either classical or prototype theory alone and, further, that they treat much of the empirical evidence which purports to concern conceptual cores as pertaining to identification procedures instead.

Having considered the three major psychological approaches to concepts and finding them unable to satisfy Kripke's considerations, we turn to consider, in more detail, the nature of the Kripke-Putnam position on natural kinds. Some consequences of this position we find unsatisfactory. One is the acceptance of essentialism; the other is the separation of epistemology from semantics. We re-iterate the position we adopted in the previous chapter, that of viewing the Kripke-Putnam position as an inadvertent *reductio* on the conception of analyticity which it presupposes. We then turn to a brief presentation of an alternative conception of natural kinds, one which we claim maintains both philosophical and psychological respectability. Finally, we suggest that this may offer the prospect of a partial reconciliation between psychological and contemporary philosophical views.

6.1 Classical Theory

In this section we detail a number of approaches to concepts which presuppose what has become known as the classical theory of concepts. Quite simply, this theory assumes that

a concept specifies necessary and sufficient conditions for membership of a category. So, for example, the concept *apple* would specify necessary and sufficient conditions for an entity to be an apple. Accordingly, every instance which falls under a classically defined concept should possess certain properties which are common to all those instances. Before we consider some of the problems of this view of concepts, let us examine a few approaches which assume this classical approach.

Hull's (1920) study involved the presentation of different Chinese ideographs to subjects. He presented each subject with twelve ideographs, each being paired with a nonsense name. Having learnt the pairing of names and ideographs, each subject would successive sets of twelve ideographs, each of these being paired with the same nonsense names as before. The design of Hull's ideographs was intriguing for, common to each ideograph bearing the same name, was a common element. And, as Hull found, successful performance in a test phase was apparently determined by memory of this common element rather than memory of the specific ideographs. That is, there was transference from one learning set to the next, the conjecture being that this was due to the fact that what subjects were exhibiting was learning of the common elements of these learning sets. So, this early investigation of concept acquisition seems to confirm the view that the process of acquiring a concept is the process of determining consciously or otherwise the common elements of the corresponding category.

A similar investigation by Vygotsky (1934) also presupposes the efficacy of this classical view of concepts. Vygotsky's sorting task involved twenty-two blocks of varying shapes, dimensions and colours. Each block has on its underside one of four nonsense names. These names correspond to four categories: large thick blocks, large thin blocks, small thick blocks and small thin blocks. The experimenter turns over one block revealing its name and the subject is asked to pick out blocks of the same sort. The experimenter corrects the subject by revealing the names of blocks until the sorting can be achieved correctly. Now many children exhibit failures to sort correctly. Instead, they appear to exhibit complexive thinking and, more particularly, *chain complexes.* These are complexes where a child in, say, the block sorting task, will proceed from item to item on the basis of attributes shared by two or more consecutive items but not by all of

the items. Such behaviour has typically been taken to suggest that these children have not acquired the concepts which the task requires. Again, the suggestion being that to acquire a concept is to have determined the common elements of a category.

However, it is perhaps unnecessary to dredge too deeply into the psychological literature on concepts since there are many adequate reviews. Armstrong, Gleitman & Gleitman (1983) provide a brief statement of the classical position with respect to concepts. A classical concept specifies a set of singly necessary and jointly sufficient conditions. Membership of the corresponding category is determined by satisfaction or otherwise of these conditions and, consequently, membership is "all or none". That is, either an entity is or is not a member of a category.

Now, the overwhelming problem faced by this classical view of concepts is that of specifying these conditions of membership. That is, the problem of specifying what it is that makes an entity a bird, say, or an elephant. Such conditions may be easy to state in the case of artificial concepts such as those defined for the purposes of the experiments of Hull and Vygotsky, but, in the case of more natural concepts and, perhaps, more particularly, natural kind concepts, such conditions are extremely difficult to find. It is precisely this problem which forms the initial focus for Putnam (1975) and the latter part of Kripke (1972) and it is to their arguments that we now turn.

Putnam, in *Is Semantics Possible?*, begins by considering a traditional approach to semantics of which the position articulated in Katz & Postal (1964) is a particular version. The position adopted by Katz and his co-workers is that the meanings of words may be given by a set of semantic markers. These markers have a number of properties as follows.

1. A word's meaning is characterised by a string of semantic markers.
2. Semantic markers stand for concepts.
3. Each such concept is a linguistic universal.
4. The meaning of a word may be specified recursively in that they may be specified

in terms of concepts which themselves are further specified.

5. Decomposition into semantic markers is taken to account for analyticity relations.

So, for example, we might presume the meaning of the word *bachelor* to decompose into the semantic markers corresponding to "unmarried", "adult" and "male". *Kill* might similarly decompose into "cause to die". Despite the fact that one might argue that the meanings of these words and the paraphrases in terms of semantic markers are different, there are other grounds to reject such an analysis. As far as Putnam is concerned, the main complaint is that such a theory of word meaning cannot be applied to very many words of the language and, in particular, natural kind terms.

Natural kind terms such as *lemon* and *tiger* are problematic for a theory of word meaning such as Katz & Postal's precisely because such a theory requires the decomposition of a word's meaning to account for analyticity relations. Since the analytic is equated with the necessary *a priori*, the decomposition must be in terms of conditions which are necessarily true of those entities to which the word applies. The problem for natural kind terms arises because such conditions as these are difficult to identify. Putnam, for example, starts with the following, admittedly putative, definition for *lemon*: yellow colour, tart taste, certain kind of peel, etc.

As he points out, the most striking difficulty is that natural kinds have abnormal members. There are green lemons, three-legged tigers, etc. This first putative definition is one which only applies to the normal members of the kind. So, the definition cannot be correct since it supposes that statements such as *All lemons are yellow* are analytic. The fact that such statements are not analytic, that is, not necessarily true, renders the definition an incorrect analysis of the meaning of *lemon*.

Putnam considers refinements to this definition, refinements which incorporate the notion of "natural kind" and "normal member" directly in to the heart of the definition. Even such drastic measures as these, however, do not save these definitions. As Putnam argues, and we concur, they fail to express analytic (necessary) truths. Putnam's conclusion is that it is simply mistaken to attempt to provide anything so simple as an

analytic definition to account for the complex behaviour of a natural kind word.

Kripke presents a similar argument against the view that the meaning of natural kind terms may be given by definition and, in many respects, a reformulation of his views against description theories of proper names. It will be unsurprising, then, to find in this argument an echo of the previous chapter.

Kripke starts by considering Kant's supposed claim that *Gold is a yellow metal* expresses an analytic truth. That is, a truth that can be known *a priori* and which is also a necessary truth. So, Kant's claim would appear to amount to the claim that gold could not be otherwise than yellow in colour; it could not possibly be otherwise than a metal. Now, this, as suggested by Putnam's observations, seems quite incorrect. Suppose, Kripke tells us, that there are some peculiar properties of the atmosphere such that gold appears to be yellow but that, when these atmospheric effects are absent, the true colour of gold is revealed and it happens to be blue. We would not suggest that gold does not exist, that gold had turned out to be a mythical substance. Rather, the claim would be that though we thought gold was yellow, it turns out that gold is blue. That is, for Kripke and, indeed, for Putnam, the word *gold* names a kind of substance. Though we may believe gold to have certain identifying characteristics, it may well be that the characteristics which we think of are not, in fact, true of gold at all. But this error on our part does not effect the meaning of *gold* nor our concept of gold. It is simply that the word's meaning and the corresponding concept are not to be explicated in terms of identifying characteristics which could possibly be false. The explications must be in terms of necessary truths.

So, the conclusion of this stage of the arguments of Kripke and Putnam is that a simple model of word meaning or concepts such as that presented by Katz and his coworkers is woefully inadequate in explicating the meaning of, or concepts corresponding to, natural kind terms. Such theories are nothing other than a formulation of the classical view whose programme of outlining the necessary and sufficient conditions for the use of a word has proved so unsuccessful.

In many respects, the arguments of Putnam and Kripke, in dismissing theories of word meaning which fit the classical mould we outlined earlier, are akin to the arguments that Kripke advances against the Frege-Russell position on proper names. The position that Kripke attributes to Frege and Russell is that the meaning of a proper name can be given by a simple description. So, *Aristotle*, according to this view, might mean "the teacher of Alexander". Kripke's position with respect to the possibility that the meaning of proper names may be given by description was, as we saw in the last chapter, that this is only the case to the extent that these descriptions embody necessary truths. And we can see that a similar position can be adopted with respect to a classical approach to natural kind terms. Namely, the meaning of these terms may only be given by description inasmuch as the description embodies necessary truths regarding the referents of such terms. So, though the classical approach might hold that *tiger* means "large, striped feline", the fact that such a description is not necessarily true of those entities we call tigers disproves the claim that the description explicates the meaning of *tiger*. Thus, there is a likeness not only between the positions adopted by Frege and Russell and those adopted by classical theorists, but there is also a likeness in the arguments ranged against these positions.

Just as Kripke notes, the spirit if not the detail of these simple description theories is preserved by those theories which adopt the notion of the "cluster-concept". And preserving the likeness between the philosophical views of Frege and Russell and the psychological views of the classical theorists, there is also a move towards "cluster-concept" approaches in the psychological literature. It is to one such approach, prototype theory, that we now turn.

6.2 Prototype Theory

There are other reasons, beyond the failure to find necessary truths, why the simple classical view failed as a psychological theory of concepts. One is the existence of prototype effects or graded structure.

Rosch and her coworkers successfully demonstrated that subjects do not respond in category decision tasks in a simple all-or-none fashion: on the contrary, responses are graded. On one task, Rosch (1975) presented names of members of everyday categories. So, for example, for the category furniture, subjects were presented with names such as chair, coffee table, bed, etc. Similarly with the "natural" category of vegetables. Each subject was then asked to rank these names according to the degree to which they exemplify the category. Not only did the subjects find the task natural and sensible, they produced reliable rankings. Carrots, for example, were reliably judged to be more typical of vegetables than were cucumber or sprouts. Similarly, for furniture, chairs are judged to be more typical of the category than are beds and desks.

Another study, Rosch & Mervis (1975), required subjects to generate lists of attributes for 20 names of members of categories such as furniture. So, presented with the name chair, for example, subjects would write down as many attributes of chairs as they could. What Rosch & Mervis discovered was that the typicality of an exemplar seemed to lie in direct proportion to the number of attributes it shared with other exemplars. Consequently, the more attributes in common possessed by an exemplar, the more typical that exemplar is rated. And the fewer common attributes an exemplar has, the less typical that exemplar is rated as an instance of the category.

Now, on a model of concepts which treats the membership or otherwise of a category to be a matter of the satisfaction of certain necessary and sufficient conditions, such findings as these are problematic. According to the classical model entities are either members of a category or not. And, consequently, the concept provides no means of distinguishing between the members of a category since these all satisfy the required conditions. Yet it is just such a discrimination between the members of a category that the subjects in the above experiments demonstrated. These findings may be taken as indicating either of two things: that the findings are not revealing of the structure of concepts; or that the structure of concepts is such as to effect discrimination between members. Whereas the former interpretation is one which might characterise the view of a classical theorist, we turn to the latter interpretation, one which characterises prototype theory.

Though there are many versions of prototype theory, Hampton (1990) offers the following properties as defining, at least, a standard view of prototype concepts.

1. The prototype concept specifies a set of weighted attributes and it further specifies, for each attribute, the allowable range of values.

2. Similarity of an entity to the prototype concept will be determined both by a process of matching attribute-value pairs and by the weightings attached to the attributes.

3. The typicality of an exemplar is determined by that exemplar's degree of similarity to the prototype concept.

4. The prototype concept also specifies a threshold level or criterion of similarity below which a member of the category cannot fall. That is, the threshold determines membership: if the similarity of an entity is lower than threshold, then the entity does not fall under the concept; if the similarity is higher than threshold, the entity does fall under the concept.

In the next chapter, we will expand in a little more detail on some of the specific versions of prototype theory. But, we can already see that, to describe prototype concepts such as these, we will require a significantly different language for describing attribute-values structures to that which we examined in Chapter 2. In particular, for each attribute we must specify a disjunction of possible values and, for each attribute, we must specify a weighting on that attribute. Though we will leave the details of such an enriched language to others, we can offer the following example for the prototype concept for *tiger*.

$$\left[\begin{array}{lll} 4 & \text{LEGS:} & \left\{\begin{array}{l} \textit{four} \\ \textit{three} \\ \ldots \end{array}\right\} \\ 2 & \text{COLOUR:} & \left\{\begin{array}{l} \textit{tawny} \\ \textit{orange} \\ \textit{white} \end{array}\right\} \\ 1 & \text{PATTERN:} & \left\{\begin{array}{l} \textit{striped} \\ \textit{spotted} \\ \textit{plain} \\ \ldots \end{array}\right\} \end{array}\right]$$

Now, in many respects, the position adopted by prototype theorists is similar to that adopted by Searle and Strawson with respect to proper names. Indeed, we can see that the brief description of prototype theory that Hampton gives fits precisely the definition Kripke gives of a "cluster-concept" approach. Kripke's requirement that to each designating expression there corresponds some cluster of properties, is satisfied by prototype theory's claim that to each attribute of a concept there corresponds a set of possible values. The fact that the cluster of attribute-values is believed to pick out the category members falling under that concept, meets Kripke's requirement that the cluster should uniquely pick out the entities to which they apply. Whether an entity falls under a concept is determined, as Kripke might have it, by a "vote" between the cluster of attribute-values. And, should the votes not reach criterion, then that entity is said not to fall under the concept. Finally, we should note that prototype theorists, while perhaps not being explicit on this point, nonetheless require there to be a set of necessary and sufficient conditions for the application of the prototype concept. That is, whether or not the cluster is satisfied by an entity is an "all-or-none" affair. To put it another way, there are necessary and sufficient conditions on an entity satisfying the prototype concept. It is simply that such conditions are generally disjunctive and more complex than those stipulated by classical theories of concepts.

So, according to Kripke's definition of cluster-concept approaches, prototype theory is just such an approach. It would hardly be surprising, then, to find that prototype theory falls foul of Kripke's main consideration, necessity. However, there are differences between the attempt of prototype theory to express necessary truths and that of the

cluster-concept accounts of proper names. So, we shall now consider in a little more detail what we would require of prototype theory in order that it express necessary truths.

In order that the sense of a proper name express a necessary truth, it must either express the essential properties of the individual who bears that name or it must express all the possible properties of that individual. The sense of *Aristotle*, for example, could either specify his essential properties or it could disjunctively specify his possible properties. So, if *the teacher of Alexander* does not express a necessary truth, we might regard the sense of *Aristotle* as specifying the disjunction *the teacher of Alexander or the author of De Anima or* ... and so on, presumably indefinitely. What Kripke argued in *Naming and Necessity* was that such a move failed to capture the missing quality of necessity. And his proposal was that to capture the notion of necessity, a sense or concept would have to express the essential properties of the entities to which it applies.

It is similar, though not the same, in the case of natural kind terms. In order that the sense of a natural kind term expresses necessary truths about that kind, it must also either express the essential properties of the kind or it must express all the possible properties of the kind. Indeed, it is on this last option that prototype theory settles. So, the task for a prototype theory of concepts, then, is to express necessary truths concerning those entities which fall under the concept. And it attempts this task via the disjunction of contingent truths. So, as part of the descriptions of concepts and of the senses of natural kind terms, we find, as Hampton suggests, a specification of the range of possible values which each attribute may assume. Let us consider an example.

Smith & Osherson (1984), as we shall see in the next chapter, also assume that each attribute of a concept is associated with a range of possible values. So, the SHAPE attribute of *apple* is associated with the open-ended list of values, *round*, *square* and *cylindrical.* And, since these are not the only *possible* shapes that an apple may have, the actual list of values would presumably be much larger. Indeed, it is difficult to imagine what shapes would not be possible for an apple to assume, as the arguments of Putnam testify. For the shape of an apple, its colour, texture, taste and so on, are

surely prime examples of its contingent properties. The question of interest is whether or not one can capture necessity by specifying contingent possibilities. Whether one can specify a necessary truth concerning the shapes of apples by specifying each of the possible shapes.

A trivial way in which one might attempt to capture necessity is by introducing tautologies. So, for example, we might say of the concept *apple* that it specifies the shapes either round or not round. Clearly, such a disjunction will be true of every apple and, moreover, necessarily true. That is, it could not possibly be the case that an apple could fail to be either round in shape or not round in shape. However, such a trivial solution is also unsatisfactory for it fails to respect the fact that the concept of *apple* should apply to all apples and only to apples. It is just such a discrimination that the concept is hypothesised to effect. The property of being either round in shape or not round in shape is one that cannot discriminate apples from non-apples precisely because the property is true of all entities.

Another means of capturing necessity via the disjunction of contingent properties is that which we might call "exhaustive disjunction". Rather than express logical tautologies as above, this method would require us to specify the disjunction of all the possibilities inherent in the model. So, for example, if we admit of just five possible shapes, *round*, *square*, *cylindrical*, *rhombic* and *toroidal*, and assuming that all these shapes are contingent possibilities for apples, then the concept *apple* would specify all these possibilities disjunctively. Whether, in principle, this strategy will work, that is, whether it is possible to express a necessary truth concerning shapes by expressing a disjunction of possible shapes, seems open to question. It would, after all, require that there be some finite limit to the types of shape an entity could possibly assume. But, even assuming that these possibilities are finite in number, the problem remains that such a concept, while expressing necessary truths will fail to express sufficient ones. That is, a specification of the shapes that an apple may possibly have cannot serve to discriminate apples from non-apples. The disjunction of possible shapes will be satisfied by the majority of entities, not simply apples.

A similar point can be made with respect to the proposals of Searle and Strawson. These require that the referent of, say, *Aristotle*, satisfies some certain proportion of statements which are true of Aristotle. But, it certainly seems to be possible that, in other worlds if not in this one, this proportion might be satisfied by someone other than Aristotle. Just as it seems that it cannot be a *necessary* truth that Aristotle satisfies this certain proportion of statements, so it seems hardly credible that satisfaction of these same statements will provide sufficient grounds for being the referent of *Aristotle*.

The point, then, appears to be that though, in principle, it may be possible to express necessary truths in terms of the disjunction of certain contingent truths, such a move will fail to capture the fact that satisfaction of the content of a concept or sense is deemed to be *sufficient* grounds for the application of that concept or the corresponding word. The fact that prototype theory requires the specification of possible values for each attribute, renders the specification equivalent to a tautology, incapable of discriminating between instances of the concept and non-instances.[1]

Perhaps the main issue of this section is whether the possibility of specifying values disjunctively is the appropriate means of capturing the fact that concepts can apply to entities with various properties. Whereas it is presumably the case that certain disjunctions must be specified as such, the question is whether this strategy will work in general. As we have seen, the considerations of Kripke suggest that the strategy cannot be successful. To put it in Kripke's terms, prototype theory and theories like it have abandoned the letter of classical theory while retaining its spirit. And, as such, they are just as prone to the difficulties of divining necessary and sufficient conditions.

6.3 Hybrid Theories

In the previous section we indicated that there are, in general, two ways of regarding the existence of prototype effects. One can argue that they are not directly revealing

[1] It may, of course, be argued that prototype concepts do specify the appropriate sufficient conditions for a concept to apply, and that they do this by introducing a weighting on possible values. In Chapter 7 we will adduce arguments against the efficacy of this particular approach.

of conceptual structure; or that concepts effect discrimination between instances of the concept much as prototype concepts do. In the psychological literature the first interpretation is relatively popular and it is this that lies behind the so-called binary or hybrid views.

In their 1981 paper, Osherson & Smith, offer a characterisation of prototype theory in terms of fuzzy set theory (Zadeh, 1965). Having failed to account, in terms of prototype concepts, for many of the typicality effects associated with concept combinations (such as *pet fish*), they conclude that there are three alternative ways of viewing prototype theory. One possibility is that what is at fault is not prototype theory but their formalisation of it in terms of fuzzy set theory. Another possibility is that prototype theory is inconsistent in some way and should be disregarded. Osherson & Smith, however, present good arguments for rejecting both these possibilities and they opt, instead, for a hybrid theory, one in which prototype theory plays only a partial role in a complete theory of concepts.

According to their view, prototype theory is concerned only with a limited aspect of concepts, that to do with a concept's *identification procedure*. Such identification procedures are taken to specify the kind of information used to make "rapid decisions about membership" (p. 57) and are to be distinguished from concept *cores*. The core of a concept plays a role in explicating that concept's relations to other concepts, and to thoughts. So, concept cores, then, are implicated in determining the facts of concept combinations and also of the truth conditions of thoughts.

The problems with a hybrid view such as this are not insubstantial. Firstly, there is the major problem of specifying the content of a concept's core. Osherson & Smith assume that some traditional theory of concepts might correctly characterise concept cores. Armstrong, Gleitman & Gleitman (1983) also assume that a theory like classical theory must specify the content of conceptual cores. Unfortunately, this theory of concepts, classical theory, is also one that has been widely rejected and for the simple reason that, in the case of natural kind terms in particular, the contents of such concepts are nearly impossible to specify. The constraint that the concept should specify necessary

and sufficient conditions, if applied to concept cores, renders the hybrid view just as unsuccessful an elucidation of concepts as was classical theory. And if this constraint is not to be applied to concept cores, then one must be left to wonder what principled distinction can be drawn between a concept's core and its identification procedures. For, after all, if both these specify contingent properties of instances of the concept, then on what grounds do we partition this content between core and identification procedure?

Secondly, and perhaps more importantly, the proposal in favour of the core/identification procedures distinction amounts to a relegation of the psychological. That is, it relegates the psychological significance of prototypes to nothing more than a function of identification. However, as Lakoff points out, the thrust of Rosch's early work was to do with thinking and reasoning. Prototypes are not implicated merely in identification but also in "making inferences, doing calculations, making approximations, planning, comparing, making judgements, and so on — as well as in defining categories, extending them, and characterising relations among subcategories. Prototypes do a great deal of the real work of the mind, and have a wide use in rational processes" (Lakoff, 1987; p. 96). The point being that prototypes are implicated in thought and the attempt to relegate them to identification procedures must be considered to be unsatisfactory from the point of view of a study of thought. This, then, is our view of hybrid theories.

6.4 Necessity and Counterfactual Conditionals

Now, as we have mentioned above, the main argument against all these expositions of concepts and of sense is that provided by both Kripke and Putnam. Their focus is on necessity and the possibility of obtaining analytic definitions for natural kind terms. Just as we saw in the last chapter the logic of the argument is straightforward. If a word has a meaning or sense then there will be some statement, expressing an analytic truth, whose truth will depend solely on the meaning of the words in the statement. So, for example, assuming *bachelor* to have the meaning "unmarried adult male", the statement *bachelors are unmarried adult males* should be true simply in virtue of its meaning, that is, analytically true.

In the case of natural kind terms, such meanings are difficult to find as, indeed, are any statements which are deemed to be analytic in virtue of the meaning or sense of a natural kind term. The difficulty arises because analytic truths are also deemed to be ones whose truth is necessary and knowable *a priori*. The task, then, is to find some statement which both defines a natural kind term and is at the same necessarily true and knowable *a priori*. The arguments of Kripke and Putnam are such as to suggest that this task is, for all practical purposes, beyond us, and their conclusion is that the meanings of natural kind terms cannot be given by a simple analytic definition or a simple Fregean sense. The main argument for this conclusion comes from a consideration of counterfactual conditionals.

Suppose that *tiger* means "large, striped, carnivorous, quadrupedal feline". The problem with this particular putative simple definition is that there are counterfactual circumstances in which a tiger may not have these properties. Some tigers may lose one of their legs becoming three-legged as opposed to quadrupedal. Were the above definition to express a necessary truth, however, to say that a tiger had three legs would simply be a contradiction. The fact that it seems not to be a contradiction, indicates that the definition does not exemplfy the meaning of *tiger*. Similar counterfactuals can be imagined. Suppose certain tigers became mysteriously infected and never developed stripes. These beasts would still, seemingly, remain tigers, yet according to the definition they would not. And similarly with virtually all the other properties of the definition.

Arguments such as these, then, are sufficient to convince Kripke that natural kind terms, just as proper names, do not have Fregean senses. Putnam advances similar arguments and reveals another aspect to this position on natural kinds: essentialism.

> A natural kind *term* is a term that plays a special kind of role. If I describe something as a *lemon* or as an *acid*, I indicate that it is likely to have certain characteristics ...; but I also indicate that the presence of those characteristics, if they are present, is likely to be accounted for by some 'essential nature' which the thing shares with other members of the natural kind. What the essential nature is is not a matter of language analysis but of scientific theory construction;... Thus it is tempting to say that a natural kind term is simply a term that plays a certain kind of role in scientific or pre-scientific theory: the role, roughly, of pointing to common 'essen-

> tial features' or 'mechanisms' beyond and below the obvious 'distinguishing characteristics'.
>
> *(Putnam, 1975; pp. 140–141)*

So, in many respects, what Kripke and Putnam have in mind is a theory of natural kind terms which play a role analogous to proper names in that they name natural kinds, each member of the kind sharing the essence of the kind. Indeed, Kripke suggests that natural kind terms are rigid designators just as he assumes proper names to be. Now this recourse to the notion of essential properties is a vital aspect of the Kripke-Putnam position. Without it there can be no grounds for claiming something is a member of a certain kind. That is, according to Kripke and Putnam it is right to think that an entity is, say a lemon, because it possesses certain properties common to lemons, it is just that these are properties that are not obvious, perhaps ones revealed in chromosomal structure. So it is, then, that Putnam in *The Meaning of Meaning* argues that uses of *water* rely on the presupposition that the substance referred to is the *same*, in Putnam's sense of *same*, as the stuff that is referred to as *water* in the linguistic community in which these uses are embedded. The position, then, relies on the presumption of essence: on there being a fact of the matter concerning whether this bit of stuff is the *same* as some other bit of stuff.

One of the consequences of the Kripke-Putnam position is a failure to explain how it is that language users may use natural kind terms and exploit their uses to extract and convey information. The failure arises from the fact that essential properties, which may well determine the true conditions of use of a word, cannot be said to determine the way in which the word is actually used. For not only do people undoubtedly use such terms to describe entities that do not have the appropriate essential properties, but all their uses of the word must also be made on the basis of ignorance. Ignorance of whether the entity in question has the requisite essence. Since whether or not a given entity possesses certain essential properties is generally an unknown, the actual use of words must be based on properties other than essential ones. And, as we suggested in Chapter 1, it is these properties, the properties on which *uses* are made, that invest such uses with informational content. So, if we are to have a theory about how it is

that people decide to use the word *lemon* and how it is that some uses can convey information such as "yellow, oval-shaped fruit", we will need to appeal to something other than essential properties. We will need, in particular, to appeal to properties of which language users may have knowledge.

In this there is also an echo of the distinction that we drew in Chapter 1. Here we distinguished between whether an entity can be *said* to be an X, that is, whether the word *X* might apply to it, and whether that entity can be said to be an X. Now the arguments of Kripke and Putnam are such as to run these two questions together and it is precisely because of the presumption that natural kinds have essential properties together with the assumption that natural kind terms are like proper names. That is, they name natural kinds. Consequently, the only answer to the question of when an entity can be *said* to be an X, where *X* is a natural kind word, is the answer, when that an entity *is* an X.

So, we have good reasons to view the conclusions of Kripke and Putnam with some scepticism and, as we suggested in the last chapter, our position is one of viewing the argument as an inadvertent *reductio* on the conception of analyticity which it presumes. To re-iterate, our position is as follows. Kripke and Putnam presume that analytic statements are ones whose truth is both necessary and knowable *a priori*. The failure to find any necessary truths underlying the behaviour of natural kind terms leads them to suppose the existence of essential properties and to suppose, further, that natural kind terms do not possess simple Fregean senses. There are reasons, however, to regard this conclusion as unsatisfactory. Firstly, we are led to postulate properties which are necessarily transcendent of our knowledge. At the current time, for example, scientists may well conclude that an entity is a lemon because of its chromosomal structure. But it is consistent with the Kripke-Putnam argument that these scientists may well be wrong. That in one hundred years time these chromosomal properties turn out to be more like obvious "distinguishing characteristics" and that some other property appears to be more essential to lemonhood is, however unlikely, a possibility. Now, according to the arguments of Kripke and Putnam, were this to be the case then the meaning of *lemon* would not have changed. What would have changed would be our knowledge of

the meaning of the word *lemon*. One problem with this view is that it follows that it can never be said, with any certainty, that a given language user either knows or has grasped the meaning of a word. For it is always logically possible that what she knows are more akin to distinguishing properties than to essential ones. Another problem with this view is, of course, the explanation of what *does* change in hypothetical cases such as we have considered. Concomitant with a change in our knowledge of the meaning of a word, comes a change in our *use* of that word and, consequently, a change in the information that such uses can convey. For our enterprise of explicating a theory of word meaning in terms that account for the information associated with the *uses* of words, the Kripke-Putnam position, which argues that changes of word *use* do not imply a change of meaning, seems inappropriate.

Their conclusion, however, may be viewed as so unsatisfactory for a psychologically acceptable theory of semantics, one that attempts to explain the actual uses of language, that we may treat the argument as an inadvertent *reductio*. Indeed, it seems that this possibility is precisely what Putnam foresaw in naming his article *Is Semantics Possible?* It is the assertion that analytic statements be ones whose truth is necessary that leads to this apparent *reductio*. Consequently, our position is that the analytic may be something other than the necessary *a priori*. In particular, as we suggested in the previous chapter, we will pursue the line that it is the contingent *a priori*.

6.5 The Relational View

By taking contingent *a priori* truths to be constitutive of meaning, we can attempt to rescue description theories of natural kind terms. Let us suppose that the following description gives one set of appropriate conditions for uses of the word *tiger*. Further, let us assume that the following description describes the default sense of *tiger*.

GENUS:	*feline*
LEGS:	*four*
COLOUR:	*tawny*
PATTERN:	*striped*
DIET:	*carnivore*

Such a description is reminiscent of the kinds of description that the classical theory attempted to associate with senses and concepts. The difference, however, is that the standard argument against such formulations, namely that they do not express necessary truths, is no longer considered valid. And this is for the simple reason that we do not require these descriptions to express necessary truths. It is enough that if someone either knows or has grasped the meaning of *tiger*, that she either knows or has otherwise grasped this description, and, further, can manifest their grasping. So, though the precise nature of the description is clearly open to empirical investigation, the quality of the description, whether it expresses a necessary or contingent truth is no longer at issue.

However, in some sense, this is misleading. For there is a real problem with descriptions such as that above. They simply do not apply to all tigers, and, indeed, they may apply to non-tigers. The latter consideration should not worry us unduly since we have already noted that *tiger* may apply to statues and the like. But the former consideration does provide grounds for concern. And the concern is, quite simply, that there are other circumstances, possibly counterfactual circumstances, in which tigers do not have the properties indicated above.

We indicated earlier (Chapter 4) how we intend accommodating such circumstances as these. Let us consider the circumstances under which a tiger has only three legs. Such a beast might well satisfy the description given below, though it will not satisfy that given above.

$$\begin{bmatrix} \text{GENUS:} & \textit{feline} \\ \text{LEGS:} & \textit{three} \\ \text{COLOUR:} & \textit{tawny} \\ \text{PATTERN:} & \textit{striped} \\ \text{DIET:} & \textit{carnivore} \end{bmatrix}$$

The problem is how to account for the fact that a language user may still call this creature a tiger, despite it not satisfying the description which that language user associates with the word *tiger*. In Chapter 4 we attempted to solve this problem by considering the description that would result from combining concepts. In this case, let us consider combining the concepts *tiger* and, say, *amputee*. Now the concept for *amputee*, let us suppose, informs us that an entity which had x limbs now has $x - 1$ limbs. Though our attribute-value language cannot represent this relation perspicuously, we will write it as follows.

$$\text{AMPUTEE}\left(\begin{bmatrix} *: & * \\ \text{LEGS:} & x \end{bmatrix} \begin{bmatrix} *: & * \\ \text{LEGS:} & x-1 \end{bmatrix}\right)$$

That is, the concept for *amputee* expresses a relation between kinds of description. Essentially, it says that, given any description of the type indicated by

$$\text{LEGS: } \textit{four}$$

it will return the very same description except for one change. Namely, that the above description will be replaced by the following:

$$\text{LEGS: } \textit{three}$$

Combining the description given for the concept *tiger* with that given for the concept *amputee* results in the following description.

GENUS:	*feline*
LEGS:	*three*
COLOUR:	*tawny*
PATTERN:	*striped*
DIET:	*carnivore*

This description *is* satisfied by our three-legged tiger, though, presumably, combining the concept *tiger* with concepts other than *amputee* would result in similar descriptions. Likewise, one can imagine what may happen in the cases of genetically abnormal tigers, ones whose stripes have faded, even ones whose four-leggedness is simply the result of an optical illusion. In each case we can assume that the appropriate description may be arrived at by a suitable process of concept combination. The concepts *genetically abnormal*, *faded*, and *optical illusion*, for example, are all ones which may be implicated. Indeed, in this regard our position is not so distant from Putnam's. Just as the importance of considering our theories of what may happen to natural kinds is obvious to him, so it is to us. All the theories which he explicitly raises, it seems, could be implicated in the process of generating appropriate descriptions or senses.

Consequently, though we have not provided exhaustive analyses of multiple examples, our position is that it is by these means that we can argue that description theories of natural kind terms may be rescued.

6.6 A Reconciliation

In the earlier sections of this chapter we attempted to highlight two trends in the literatures of contemporary psychology and analytic philosophy. Firstly, we attempted to show that the same simple analytic conception pervaded both, leading to classical theory in the psychological literature on concepts and the positions of Frege and Russell with regard to proper names. The first trend, then, was the trend away from these views to those defined by cluster-concept approaches. In psychology, we see this in prototype theory and, in contemporary analytic philosophy, we see this in the rejection of the Frege-Russell position by philosophers such as Searle and Strawson. The second

trend, however, is a move away from description theories altogether and, though we see this in the arguments of Kripke and Putnam, there are few comparable arguments in the psychological literature. Indeed, the prevailing view of concepts appears to be that these are best described by some version of prototype theory. So, the second trend is a divergence of psychological and philosophical views. The current framework, one which concentrates on the informational content associated with word uses, has attempted to attain degrees of both philosophical and psychological respectability. In order to show this, let us consider in more detail the relation between the view I have proposed and that which Putnam proposes in his attempt to argue that semantics is, despite its difficulties, still possible.

Putnam's solution accords considerable significance to the fact that the meanings of (certain) words may be conveyed with remarkable ease.

> The fact that one *can* acquire the use of an indefinite number of new words, and on the basis of simple 'statements of what they mean', is an amazing fact: it is *the* fact, I repeat, on which semantic theory rests.
>
> *(Putnam, 1975; p. 149)*

So, any theory of word meaning should presumably respect this central fact and it is this that Putnam's notion of core fact is intended to do. For Putnam, associated with each natural kind term is a stereotype: "the associated idea of the characteristics of a normal member of the kind". Now, to convey the meaning of a term is to convey the core facts associated with that term which are, according to Putnam, the stereotype and the extension of the term. However, as Putnam goes on, the extension need not always be given since the hearer may always consult an expert on the matter of the appropriate extension for a given term. In certain circumstances, though, where the hearer may attach the kind word to the wrong natural kind, extra help must be given. That is, it is necessary to communicate some means of getting the extension right, perhaps by giving a sample of the correct extension, just as one might give a colour sample.

Now, it is clear that much of what Putnam has said is consistent with our account. Firstly, the fact that we may associate a default description with the concept of a

natural kind term respects Putnam's considerations in favour of the notion of stereotype. And, indeed, it seems that, in many ways, Putnam's conclusion is a vindication of our position. Particularly, in the way that we treat the role of theories in natural kinds. It is, after all, precisely those explanations that Putnam gives as to why a particular member of a kind has exceptional properties, that play the role, in our framework, of determining whether that entity can be described by a use of the corresponding natural kind term. That is, when Putnam tells us of the possibility of a gas descending into the atmosphere, turning all lemons blue, he has provided precisely the kind of "explanation" that, according to our framework, language users will employ in determining whether a blue, oval-shaped fruit can be called a lemon.

Where we disagree with Putnam is in his claim that descriptions such as those we have given do not explicate the *meanings* of natural kind terms. But, as we have argued, this depends upon one's prior conception of what meaning relations should be like. In particular, if one regards meaning relations as being contingent then those descriptions we have given can be thought of as explicating meaning. It is only the insistence that meaning relations should have the property of necessity that suggests the contrary. Indeed, it is the insistence that a natural kind term should act as a name, a rigid designator, of a kind that disallows the possibility of explicating an account of natural kind terms in terms of descriptions. Yet it is just this kind of account that is being sought in the psychological literature where the concern is less with the question of whether a given entity *is* a such-and-such and more with whether the entity can be *said* to be a such-and-such, more with whether the entity falls under the concept *such-and-such*. It is in this spirit, the spirit of moving toward a reconciliation of contemporary psychological and philosophical views, that we propose the possibility of rescuing description theories of natural kind terms: a possibility raised by construing the uses of these words as betraying multiple meaning relations, each having the properties of being both contingent and *a priori* knowable relations.

Chapter 7

Reinterpreting Prototype Theory

> Instead, the concepts that people use are constructed in working memory from knowledge in long-term memory by a process sensitive to context and recent experience. Concepts in working memory may be stable under some conditions, however, concepts typically appear to vary widely as a function of goals, current context, and recent experience. Although theoretical abstractions that correspond to invariant concepts may serve some useful theoretical and empirical purposes, it may generally be more useful and more accurate to view concepts as temporary constructions in working memory.
>
> *(Barsalou, 1987; p. 135)*

In Chapter 1 we talked about the relations between the studies of word meaning and of concepts. In Chapter 4 we discussed some of the literature on concepts in order to motivate a particular view of sense generation. In this chapter, we return to a theory of concepts, namely, prototype theory, and re-examine it from the perspective of the relational view we have developed.

As we suggested in the previous chapter, prototype theory stands little obvious chance of withstanding the force of a Kripkean attack. In evaluating the claim that senses must be explicable in terms of necessary truths, there are two options for prototype theory. One is to conclude that prototypes *do* express necessary truths by means of the disjunction of contingent ones. The other is to claim that prototypes do not describe senses. This last option, it seems, is the one that Kripke would have us take. But, as we

have seen, psychology seems to be of the opinion that prototypes do describe senses. At least, it is *prototypes* that are deemed to be implicated in the compositional semantics of sentences and phrases. This impasse between philosophical and psychological views is of exactly the same nature as that between Kripke's position on proper names and the position of the description theorists. Just as we saw in Chapter 6, a way of restoring confidence in the psychological claim that senses may be described by concepts is to treat the Kripkean argument as a *reductio* on the conception of analyticity which it presumes.

Thus, one goal of this chapter is to spell out a plausible alternative to the traditional analytic view of concepts. This we attempt in section 7.3. The means by which we approach this task is via an evaluation of prototype theory. In the next section, then, we offer an outline of prototype theory and classify it in terms of the choices of Chapter 2. In section 7.2 we detail a number of problems with prototype theory. Some of these are directly suggestive of an alternative, namely the Relational View, which we consider in section 7.3 though we now rename it the Family of Constraints (or FoC) view. After a brief re-iteration of this view's main points, we turn to re-interpreting prototype theory by means of a detailed comparison with the FoC view. Among the results of this comparison are the suggestions that the FoC view is better placed to interpret both Barsalou's claim that concepts are simply convenient fictions and Wittgenstein's observations and comments concerning family resemblance. Indeed, it is this fact that suggests the new name for the view presented in Chapter 4.

The motivation for this comparison between prototype theory and the FoC view is twofold. Firstly, we hope that the claim concerning prototype theory is not simply that it is flawed in certain respects, but that these flaws appear because it may be attempting to describe problems at a different level from the FoC view. That is, we hope to make sense of the flaws in prototype theory. Secondly, the comparison is motivated by the fact that the assumptions of the FoC view are such as to undermine the motivation for prototype theory. Let us consider this second reason in a little more detail.

One way of examining the psychological enterprise of viewing the senses of words in

terms of prototype concepts, is as an explicit attempt to capture the *necessity* of such senses. For example, Smith & Osherson's (1984) model assumes that, for each attribute relevant to a category, the prototype concept specifies the set of possible values. That is, the set of values that it is possible for the instances of the category to assume. Specifying all such possibilities would, indeed, amount to specifying all counterfactual possibilities and such a specification, in disjunctive terms, would clearly be necessary of every instance of the category.

Now what is of interest for our purposes is the consequence of adopting our alternative conception of analyticity, for this does not require a sense to express necessary truths. Indeed, they may express truths which are only contingent. So, such a conception of analyticity obviates the motivation for this move by prototype theorists in favour of disjunctive properties, though, it does not, of course, obviate the empirical observations which motivate the notion of prototypes. And, in this, lies the interest. For, if the need for a prototype treatment of senses is obviated, how are we to interpret prototype theory? How much, then, of prototype theory is in accordance with our FoC view and how much not? Answering these questions, then, forms the other major focus of the present chapter.

We begin, then, by outlining the basic features of prototype theory and two recent models of prototype concepts.

7.1 Classifying Prototype Theory

The choices in Chapter 2 principally concern word meanings and how they might be mentally represented in terms of the components of some "mental lexicon". In section 1.3 we hinted at the close relation between concepts and word meanings and, indeed, in Chapter 6 this relationship was further specified: our discussion has illustrated that, within the psychological literature, the mental representations of word meanings are taken to be concepts. That is, (at least, a subset of) concepts are taken to be components of some "mental lexicon". In this section we will, therefore, be concerned to elucidate

precisely what claims are made by a particular theory of concepts, namely prototype theory, in terms of the choices for theories of word meaning we have already outlined. Before we consider these choices, let us outline the theory under consideration, prototype theory.

7.1.1 Prototype Theory

Prototype theory is an umbrella term for a large number of different approaches to concepts and word meaning. Though a number of these are quite recent, the history of prototype theory is relatively long, at least in terms of contemporary cognitive psychology, and many of its roots can be traced to the debate concerning the possible relativity of thought to language. Our goal in this section is far more modest than a detailed history of prototype theory would require, and it is, quite simply, to outline the *current* conception of prototype theory, the conception which guides a considerable amount of contemporary research into concepts.

As we mentioned in the previous chapter, Hampton (1990) offers a number of properties which may be taken as defining the standard view of prototype concepts. Namely, the following four conditions. The prototype concept specifies a set of weighted attributes and it further specifies, for each attribute, the allowable range of values. Similarity of an entity to the prototype concept is determined both by a process of matching attribute-value pairs and by the weightings attached to the attributes. The typicality of an exemplar is determined by that exemplar's degree of similarity to the prototype concept. The prototype concept also specifies a threshold level or criterion of similarity below which a member of the category cannot fall. That is, the threshold determines membership: if the similarity of an entity is lower than threshold, then the entity does not fall under the concept; if the similarity is higher than threshold, the entity does fall under the concept.

Theories which match these conditions are not too difficult to imagine. Indeed, the conditions above are highly reminiscent of those we enumerated in Chapter 5 in defining

"cluster-concept" approaches to proper names. But to place a little flesh on these meagre bones, let us reflect on a couple of the more explicit versions of prototype theory. These are firstly, the "Knowledge Representation Model" of Cohen & Murphy (1984) and secondly, the "Selective Modification Model approach of Smith et. al. (1988).

Cohen & Murphy's model has four basic features.

1. Concepts have *roles* which are filled by feature lists or *values.*
2. Concepts are organised hierarchically so that certain roles and values may be *inherited*, though only by default.
3. Roles are *restricted* so as to take only particular values.
4. The instances or tokens of concepts are also concept-like descriptions that are subsumed by the concept.

Further to these points, Cohen & Murphy also assume that the possible values for a given role are *ordered* according to typicality. They also assume that the subordinates of a given concept are ordered according to typicality. So, for example, the subordinates of the concept *bird* would be ordered ... *robin* ... *chicken* ... *penguin* .... Similarly, the values are ordered. The concept *bird*, for example, has a MOVES role whose possible values are ordered ... *flies* ... *walks* ... *swims* .... This is portrayed in Table 7.1.

The other version of prototype theory we will consider is Smith & Osherson's (1984) "representational" model. In their later paper, Smith et. al. (1988) propose a "Selective Modification Model" which, while it is similar in spirit, is not identical to the "Representational" model they proposed in Smith & Osherson (1984). Hence, the outline is based on the formal presentation of the later paper, though some of the comments I draw on are to be found in the earlier paper. Wherever possible, I will try to indicate the precise source.

Smith & Osherson assume a representation for prototype concepts that embodies the following aspects.

ISA:	*animal*, ...
SPECIES:	*robin*, *chicken*, *penguin*, ...
EXTERIOR:	*wings*, *beak*, ...
MOVES:	*flies*, *walks*, *swims*, ...

Table 7.1: A Schematic for Cohen & Murphy's (1984) Knowledge Representation Model: the Prototype Concept for *Bird*.

1. The representation possesses a set of relevant *attributes* for the concept.
2. For each attribute, the representation specifies a set of *possible values*, or features, that instances of the concept can assume.
3. The representation specifies a *weighting* ("votes"), or salience measure, for each possible value of each attribute.
4. Each attribute is itself weighted, this being termed that attribute's *diagnosticity*.

Smith & Osherson assume that objects may be described in a similar manner and that the process of determining whether or not a given object is an instance of the concept is a relatively complex process of attribute-matching. It goes as follows.

An object will typically receive "votes" for only one possible value of each attribute. A prototype concept will typically receive votes for many such possible values. The similarity of an object to a prototype concept is determined by a version of Tversky's (1977) "contrast" rule. That is, for each attribute, the number of votes which match for both the object and the prototype concept are summed and subtracted from this are the number of votes for which no match is achieved. For example, assume that the

1	COLOUR:	*red*	25
		green	5
		brown	
		...	
0.5	SHAPE:	*round*	15
		square	
		cylindrical	5
		...	
0.25	TEXTURE:	*smooth*	25
		rough	5
		bumpy	
		...	

Table 7.2: A Schematic for Smith & Osherson's (1988) Selective Modification Model: the Prototype Concept for *Apple*.

prototype concept has the following weighted values for the attribute COLOUR: *red* 25; *green* 5; *brown* 0. Assume that the object has the following for its COLOUR attribute: *red* 20. Then the number of matches are 20 (on the *red* value) and the number of mismatches are 10 (5 from the *red* value, 5 from *green* and 0 from *brown*). So the overall "matching" sum is $20 - (5 + 5 + 0) = 10$. This number is then multiplied by the diagnosticity of the *colour* attribute and the result for each attribute is summed to give an overall total of the similarity of the object to the prototype concept. One of the great virtues of Smith & Osherson's model, then, is its explicit formulation. In Table 7.2 we reproduce their description of the prototype concept for *apple*.

In the following sections we will have much to say concerning the nature of the two models we have considered and the nature of prototype theory, in general. For now, however, we will simply point out one of the ways in which these models of prototype concepts differ. Both the models of Cohen & Murphy and Smith & Osherson explicitly encode typicality differences in terms of an ordering over the values a given attribute may take. However, in each case, the precise details are distinct. Cohen & Murphy's model assumes a simple linear, that is, ordinal, ordering: the prototype concept contains lists of values where ordinal position in the list reflects typicality. Smith & Osherson's model, however, by explicitly introducing weights on values, allows the typicality of these values

to be essentially free-valued. That is, values may be listed according to the size of the corresponding weight, but typicality is determined by the size of that weight, not by ordinal position in the list. So, whatever else one may say about these models, they do at least have interesting differences. The model of Cohen & Murphy, for instance, would hold that, for a particular attribute of a particular concept, the difference in typicality between the second and third most typical values equals the difference in typicality between the least and next to least most typical values. Whether this would suggest that the difference in "birdness" between a robin and a sparrow equals that between an emu and a penguin is a matter for Cohen & Murphy. That the model of Smith & Osherson is not similarly committed is clear from the fact that their model stipulates no relation between the weightings of different values. For them, the determination of such weights is, presumably, an empirical matter.

This difference aside, the two models are very similar and both fit Hampton's desiderata concerning prototype theory. In the remaining sections we outline some of the arguments against models such as these, before engaging in a comparison of these models with the Relational View, or FoC view, outlined in Chapter 4.

7.1.2 Are Prototypes Conditional or Unconditional?

In section 1.4, we suggested that under those views that equate concepts with the mental representation of word meanings, concepts can be seen to stand in the relation of *attunement* to word meanings. However, this raises the issue of the kind of meaning relations to which concepts express attunement. In particular, whether such relations are unconditional or conditional.

As we saw in section 1.4, unconditional constraints have the property of holding in all situations. No matter what situation is under consideration it must be of a type such that the constraint is a component of that situation. Conditional constraints, on the other hand, hold only in certain types of situation. That is, not all situations must have conditional constraints as components. Now, to see whether we should conceive of pro-

totype concepts as expressing attunement to unconditional or conditional constraints, let us proceed as follows.

Firstly, let us assume that concepts express attunement to *conditional* constraints. One conditional constraint pertaining to apples, for example, might be that they have pips. Such a constraint is conditional because we can conceive of conditions, such as those relating to genetic malformation, under which things which are apples do not have pips. Yet such a constraint *is* a constraint, because it is the case that under certain circumstances knowing such-and-such is an apple can carry information to the effect that the very same such-and-such has pips. In situations, for example, where all the apples do have pips. So, let us assume, for sake of argument, that this is the constraint to which our concept of *apple* expresses attunement. We can write it as follows.

$$S_1 \Rightarrow S_2 \mid B$$

where B is the type of situation where, among others, the conditions associated with certain kinds of genetic malformation do not hold, S_1 is the type of situation in which a given entity is an apple, and S_2 is the type of situation in which that same entity has pips. Let us write this more informally as follows.

$$apple(x) \Rightarrow has\text{-}pips(x) \mid B$$

To be consistent with our discussion of prototype theory, we can formulate a related concept-like description in terms of attributes and values as follows.

$$\left[\text{CONTAINS: } \left\{ pips \right\} \right]$$

So, the claim that we are examining is, to be precise, not exactly the claim that concepts express attunement to conditional constraints, but that they express attunement *only* to conditional constraints. Now, there are several points that can be made in response to

this claim. Firstly, the above constraint applies only in situations of type B and yet this *local* nature is not traditionally a characteristic of concepts. Concepts, it is traditionally assumed, apply in all situations in which entities falling under that concept exist. Since, there are thought to be no *a priori* conditions on the *existence* of entities, that is, it cannot be said *a priori* that such-and-such thing cannot *exist* in such-and-such situation there can be no restrictions on the domain of application of the concept.[1] That is, according to traditional views we cannot claim that the concept of *apple* applies only in certain situations.

Secondly, the question must be raised as to what is involved in deciding that an apple without pips is, nonetheless, an apple. For this job is traditionally conceived as that of the concept for *apple*. Yet if the above constraint is constitutive of this concept, then it cannot deliver *any* judgement on an apple without pips. So, according to traditional views, a concept must be much richer in information than concepts like that above. It must deliver judgement on *all* entities to which it applies, not just some. That is, the concept of *apple* must apply to all apples, not just those with pips.

Such considerations as these do indeed inform traditional views of concepts and the position we develop later will be in stark opposition to these views. However, for present purposes it is sufficient to recognise that traditional theories of concepts require a concept X to apply to all entities which are X, and that, since there can be no *a priori* argument to the effect that Xs can only exist in certain types of situation, such traditional views require concepts to apply, in principle, in *all* situations. That is, traditional views of concepts are, in our parlance, seen as expressing attunement to *unconditional* constraints.

This fact can be seen in one of the oft cited claims for support for prototype theory, family resemblance. Wittgenstein's (1953) observations of family resemblance concern the failure to find any common elements underlying the category of games. A similar observation holds for our example, apples: some are pipless, others have pips. The possession of pips, then, is not a common property of apples and, in this, apples appear

[1] Of course, this is not true if we allow circularity. That is, we could say that apples do not exist in situation of type A, and define A as the type of situation in which apples do not exist.

to exhibit a family resemblance structure. In prototype theory, such facts as these are taken as evidence not that there are no common properties for certain categories, but that, for those categories, there are no common simple properties. To be explicit, the simple property *having pips* would not be common, but the complex disjunctive property *having pips or not having pips* would be common. And, in this move of explicitly introducing disjunction, one can see the goal of trying to preserve the notion that there may be common properties to the instances of a category. Indeed, one can go further and suggest that, in this way, prototype theory is attempting to establish conditions which are truly necessary of instances of the concept. Whether this genuinely reflects Wittgenstein's own views on family resemblance is, of course, open to question, but the quote with which Chapter 5 began suggests otherwise. Indeed, we will assume that it *was* Wittgenstein's point that there may be *no* truly necessary conditions on the instances of a concept, even complex, disjunctive conditions. That is, following Wittgenstein, suggesting that disjunctive properties are common properties is tantamount to playing with words. Nonetheless, interpreting prototype theory in this way, as an attempt to establish necessary and sufficient *complex* conditions on the application of a concept, makes the view that prototype concepts express attunement to unconditional constraints even more compelling.

Our conclusion, therefore, is that if we view concepts as expressing attunement to constraints then prototype concepts express attunement to unconditional constraints.

7.1.3 One or More

In Chapter 2, we discussed the logical possibility of regarding the various uses of a single word as exploiting more than one meaning relation or constraint. The counterpart of this choice for our discussion of concepts is whether or not prototype theory is committed to the view that more than just one concept X is implicated in determining whether an entity is or is not an X.

The models of Cohen & Murphy and Smith & Osherson (1984) are unequivocally com-

mitted to the view that the process of judging whether or not a given instance falls under a concept is a function of the description of that instance and of the appropriate prototype concept. For example, determining whether a given individual is an apple, depends on how we describe the properties of the individual and on the concept *apple*. So their view is that whether or not a given object is an instance of a category is determined by the *one* concept for that category. As we shall see later, this view has some unfortunate consequences with respect to a number of psychological considerations. However, for now, we will simply note that, as far as prototype theory is explicit on this matter, the determination of whether a given entity is or is not an instance of a category X is achieved by the single concept *X*.

Before concluding this section, let me make it clear that we are not suggesting that Smith & Osherson cannot account for *certain* effects of interaction among concepts. For example, having the concept of *cup* distinct from the concept of *mug*, may enable us to think of cups *as* cups. It may also, in many cases, prevent us from thinking of cups as mugs and vice versa. However, were we not to have the concept of *cup* we might well normally think of cups as mugs. Thus, a concept may prohibit other concepts from applying to entities within its domain. In this, concepts may behave similarly to words in that the presence of one word may *block* other words. For example, the word *cup* may prevent or block the use of the phrase *mug* from applying to cups. Ignorance of the word *cup*, however, allows *mug* to apply to cups.

So, we cannot, and nor do we mean to, discount the possibility that Smith & Osherson can account for various forms of interaction among concepts. However, in terms of a synchronic account of which conceptual organisation determines whether an entity is or is not an X, the assumption is that this determination is achieved by the single concept *X*. That is, not only are prototype concepts seen as expressing attunement to unconditional constraints, it is deemed that each expresses attunement to one single constraint. One unconditional constraint, then, is thought to determine the conditions of application of concepts.

7.2 Some Apparent Problems

In this section we consider only some of the more obvious problems associated with prototype theory. That is, we are not attempting to exhaustively evaluate prototype theory, but merely those aspects on which we will later compare prototype theory with the Relation View, or FoC view, developed in Chapter 4. Necessarily, some of these issues we will have already discussed in motivating the Relational View but, in most cases, the discussion here will be significantly different. Our first consideration concerns the adequacy or otherwise of prototype theory in accounting for the development of concepts.

7.2.1 Coherence

Murphy & Medin (1985) concentrate on an issue to which any theory of concepts should have an answer: conceptual coherence. Roughly, the question is as follows: what makes such-and-such group of individuals fall together under the same concept? As we have already seen, the view according to prototype theory is that individuals "fall together" because of some intrinsic similarity to one another, this similarity being determined by a prototype concept. That is, all these individuals will be similar to one another to the degree that they may share certain attributes and values. Should they share sufficient of these, then they will be judged similar, and grouped under the same concept. So similarity, being determined by attribute-value matching, determines the coherence of concepts. It is the thrust of Murphy & Medin's article that this view of conceptual coherence is woefully inadequate.

Murphy & Medin's enterprise is one of demonstrating that any approach to conceptual coherence which ignores higher level "theories" of the world will be lacking. The specific goal which is of interest given our current purposes, however, concerns whether the models of prototype concepts offered earlier can capture the coherence of concepts. One example Murphy & Medin offer concerns the concept *intoxicated.* It is quite unlikely that such a concept would incorporate the description *jumps into swimming pools fully*

clothed, yet, on certain occasions, the truth or otherwise of this description may well determine whether an individual is categorised as being intoxicated. Murphy & Medin's explanation of examples such as this is that acts of categorisation are made with reference to our knowledge of theories of the world. So, for example, in categorising someone as intoxicated, we are aided by our theories of intoxicated individuals, theories of house parties, theories of drowning and of heroism, among others. That is, categorising an individual as falling under a concept is not determined solely by that concept, but rather by that concept in conjunction with numerous relevant theories.

Let us be a little more precise about this. As we indicated earlier (section 4.1), Murphy & Medin also note the important role in categorisation played by our knowledge of "constraints about operations that are permissible" (p. 295). They offer the following example.

> Suppose that all the soda cans you have come into contact with have been 7.5cm in diameter and that all the silver dollars you have seen have been 4cm in diameter. Suppose further that you are told that some entity has a diameter of 5cm and you are asked whether it is more likely to be a soda can or a silver dollar.
>
> *(Murphy & Medin, 1985; pp. 295–296)*

Murphy & Medin suggest that one is more likely to judge the 5cm wide object as a soda can since it is an aspect of our theories of monetary systems that coins of a given denomination are of a uniform size. That is, it is not similarity as computed by some simple attribute-value matching metric that determines whether we view such an entity as a can or a coin, but rather the determination arises, as Murphy & Medin state, from "knowledge about transformations and operations associated with concepts, and this, in turn, relies heavily on our general world knowledge" (p. 296).

Murphy & Medin point out that they do not intend to offer a model of concepts but, rather, to articulate constraints on any such model. The thrust of their article is to suggest that concepts are embedded in theories. Conceptual coherence is seen as arising from the fact that a concept may be related to a number of theories of how the world

operates. So, for example, we can say that determining whether an instance falls under the concept of *lemon* does not depend on a superficial similarity metric. Instead, it may depend on our ability to infer and reason about the possible attributes of lemons. We might, for example, reason just as Putnam does. That is, we might reflect on the possibility that some gas has descended into the atmosphere and turned all lemons blue. So, from then on, determining whether an instance falls under the concept *lemon* will depend not on whether the instance is yellow or, for that matter, blue, but whether we can infer some kind of relationship between its properties and the properties of those things we know to be lemons. That is, we must work out why this thing is blue when all the other lemons we have previously come across have been yellow. Provided we can reason as to why a lemon would appear blue, then blue lemons can be categorised as lemons. Presumably, the converse is also true. Namely, if we cannot reason as to why lemons might appear blue, then we will not categorise any blue thing as a lemon. Explanations such as these, it seems, is what Murphy & Medin mean by "theories".

Coupled to their view of conceptual coherence arising from theories is a suggestion that ties it simply to the number of "explanatory links". Decorating bathrooms and eating out may, for example, not appear to form a very coherent category but, if these activities are ones which occupy one's partner, then the category gains in coherence. That is, the coherence of *decorating bathrooms and eating out* is not derivative on some superficial similarity but on what else we know that relates them, on their being explanatory links, and the coherence of a concept can be enhanced simply by adding more explanatory links. So, one could offer explanations as to why one's partner commonly engaged in decorating bathrooms and eating out, and such additional explanations would add coherence.

Needless to say, all these observations concerning conceptual coherence are problematic for the models of prototype theory we examined earlier, and those like them. We can make two points in this regard, a theoretical one and a methodological one.

The theoretical point is quite simple and, indeed, ably made by Murphy & Medin: the models of prototype theory we examined earlier are inadequate as theories of concepts.

They do not capture the important observation that there are dependencies between concepts and between the attributes and values of a concept and that these dependencies play an important role in the psychology of categorisation. Now, the most obvious response to this is one that has been made by, among others, Osherson & Smith (1981). They distinguished two aspects to concepts: a core and an identification procedure. Further, they suggest that prototype theory is best seen only as a theory of the identification procedures associated with concepts and not as their core. Since the conceptual core is intended as the locus of "those aspects of a concept that explicate its relation to other concepts, and to thoughts" to suggest that prototype theory does not express such relations is, therefore, not to argue against prototype theory at all.

However, this response seems unsatisfactory. Murphy & Medin's case of the soda can/silver dollar is a case where the categorisation asked for must be made solely on the basis of perceptual information. It is, *prima facie*, a case well-suited for prototype theory, a case which asks solely for an identification judgement. But even were Osherson & Smith able to offer good reasons why such a case does not constitute a counter-example to their version of prototype theory, there would be other theoretical reasons for not accepting the identification procedure/conceptual core distinction, reasons which we outlined in section 6.3.

The methodological problem is less easy to articulate. The representations offered in explicating the models of Cohen & Murphy and Smith & Osherson can be thought of as expressions in some language. We might, for example, think of them as expressions in some (extended) attribute-value logic (cf. Johnson, 1988). The essence of such languages, however, is precisely that they deal in *attributes* and *values*. That is, they are well equipped to express facts concerning objects. We might, for example, uncontroversially assume that our world is inhabited by individuals and that these individuals either do or do not possess universal properties, such as *being red*. However, without extending our metaphysics we cannot talk of relations between individuals, nor relations between universals. That is, in order to correctly express truisms concerning such relations, we must enrich our ontology to include relations.[2]

[2]We are using *relation* here so as to exclude one-place relations. That is, here at least, we intend

Now, what Murphy & Medin have so clearly demonstrated, is the need for a theory of concepts to be able to express facts about relations, not merely properties. In order to express some similarity between decorating bathrooms and eating out, a theory should not appeal to some intrinsic properties of these activities. Rather, it should appeal to the fact that both activities stand in the same relation, say *engages-in*, to some salient individual. What should be clear is that a formal language equipped with an ontology of attributes and values is particularly ill-suited to expressing facts about relations. So, one possible response to Murphy & Medin might be along these methodological lines, that it is the formal language which is conceptually impoverished, not prototype theory itself.

While these options are indeed open, in general they must be seen as dissatisfying. In each case, they serve to seriously weaken the claim being made, sometimes even to the point of invalidating that claim. In the case of the theoretical option of appealing to a distinction between core and identification procedures, the claim that prototypes explicate concepts is seriously undermined. Indeed, they are to be regarded as explications of how subjects recognise instances of a category, not how they think of them. That is, under this interpretation, the models we have considered say little, if anything, about thought. Appealing to the methodological point is less serious. It allows one to claim that a particular version of prototype theory *can* account for conceptual coherence in the way Murphy & Medin suggest but that the formal account fails to express this fact. However, in the absence of such an account of conceptual coherence accompanying the models we have examined, we can safely conclude that these models of prototype concepts do not and cannot adequately account for conceptual coherence.

7.2.2 Developmental Adequacy

The view of concepts which prototype theory offers is, as we have seen, that the concept X determines for *all* entities whether that entity is a member of the category of Xs. So, for example, for all trees it is our concept of *tree* which determines whether we judge

relation and *property* as disjoint.

an entity to be a tree or not. Such a view holds that any developmental changes which affect judgements on treehood must reflect changes in the concept *tree*. There appear to be good reasons to think that this view is simply wrong.

Consider, by way of an example, how it is that one can judge a tree which has been struck by lightning, nonetheless to be a tree. Firstly, there can be no suggestion that this is a peculiar skill or one that requires prior knowledge that the entity in question is a tree. The fact that, on the basis of *perceptual* evidence alone, people can judge such scarred and deformed trees to be trees nonetheless, is, I take it, not open to question. However, what *is* in question is precisely which concepts they bring to bear in making such a judgement. Prototype theory would have us believe that people simply use their concept of *tree*. Consider, for example, the case of some tree which, having been struck by lightning, is now horribly disfigured. The claim that recognising this object as a tree is a function solely of the concept *tree*, amounts to the claim that such an act of recognition is in no way dependent on other concepts such as *struck by lightning*. Now, the question must be raised, what does it mean to say that someone's recognition of Xs which have been struck by lightning is in no way dependent on their having the concept of *struck by lightning*?

One possibility, the one adopted by prototype theory, is that this claim is such as to suppose that all the possible attributes for a given concept are simply disjoint. So, for instance, we may say of apples that their colour is, among others, either red, brown or green. That is, we specify their possible attributes as a disjunction. So, in specifying the concept *tree*, we can specify the possible shapes of trees similarly, to include the disjunction *R*-shaped or *S*-shaped, where the former is the normal shape of trees and the latter is the shape of trees struck by lightning. Now, under such a view, recognising a tree that has been struck by lightning as a tree, would involve determining whether the disjunction, *R*-shaped or *S*-shaped, would be satisfied by that tree. Most importantly, it would not involve the concept of *struck by lightning*.

Now, if this is intended as a theory of all concepts, as the framework of prototype theory suggests, then it seems to be straightforwardly incorrect. Consider the concept *dog*. This

concept applies to all manner of objects which are quite dissimilar in appearance. Yet it is the case that we can recognise many dogs on the basis of perceptual evidence alone. So, for instance, the concept *dog* is satisfied by the very youngest puppies and the very mangiest mongrel. Now, a prototype model of the concept *dog* might, for example, specify number of ears disjunctively. The reason is, as Cohen & Murphy and Smith & Osherson note, due to the fact that the prototype concept *dog* must specify all *possible* values for relevant attributes. Since it is undoubtedly possible that dogs may have no ears, one ear, or two or more ears, these possibilities must be listed as the terms of a disjunction. So, for the attribute NUMBER-OF-EARS, the prototype concept *dog* might assign a high positive weight for the value *two* and a low positive weight for, say, *none*. Certainly, the disjunction of these values will be satisfied by every dog. But the problem is that this disjunction fails to express the fact that if a dog has no ears this is usually for certain reasons. We may, for example, appeal to certain facts regarding the dog's genetic make-up; we may propose the existence of some chemical imbalance in the puppies foetal environment; we may speculate that the dog has lost its ears in an accident. Indeed, there are a host of possible *explanations*, just as Murphy & Medin might suggest.

So prototype theory brings with it the suggestion that earless dogs can be categorised as dogs regardless of whether there is *any* explanation for their earlessness. The suggestion is not simply that categorisation does not depend on our *knowing* the reasons why this dog has no ears; rather, the suggestion is that categorisation does not depend on our assuming that there *is* an explanation. So, the suggestion is that categorisation treats the relation between a dog with no ears and a dog with ears as being qualitatively of the same type as that between an apple which is red and one which is green: the presence or absence of a causal relationship has no effect on the categorisation.

The previous discussion of Murphy & Medin's approach to conceptual coherence provides sufficient grounds for us to view this suggestion from prototype theory with suspicion. Indeed, it seems to be plainly false. It just seems to be the case that were there to be no possible explanation *whatever* of how a dog has no ears, then we would not regard it as a dog. Unless, of course, we changed our concept of *dog*. Similarly, it seems

that people lacking the concept of *lightning*, or some similar concept, would not judge a tree which has been struck by lightning to fall under the concept *tree*. To expand, if one is not aware that entities can be altered in such a way as to leave them scarred and deformed, then one will conclude that there is no conceivable way a tree can come to look this way.

Consider another example. Suppose that you are not in the possession of the concept *sliced*. That is, for you, there is no conceivable way in which, say, a spherical object can be transformed into a number of circular slices. Now suppose that you come across a lemon which has been so sliced. The claim that I wish to make is that you will not admit that such an entity constitutes a lemon, rather that it must be seen as a number of distinct circular objects which bear certain similarities to lemons. And this is not because of some deficit in your concept of *lemon*, rather because you don't possess the concept *sliced*.

So, the claim is that developmental changes in judgements of whether an entity is or is not an X, should be seen not merely as arising from changes in the concept *X* but also from changes in, and the acquisition of, other concepts. As we have seen this is both something which prototype theory cannot account for and something which is perfectly consonant with the approach of Murphy & Medin. However, before turning to the next possible problem with prototype theory, let us consider an objection to this claim concerning development.

It may be objected that the examples concerning lemons and trees show how the concepts of *lemon* and *tree* change on encountering more instances of the appropriate categories. So, it might be argued, when we encounter this sliced lemon, initially we don't judge it to be a lemon because we are not aware that lemons can look this way. That is, our prototype concept does not specify *sliced-shape* as one of the possible values for SHAPE. However, on being told that this sliced thing *is* a lemon, our concept of *lemon* changes: the list of possible values for the SHAPE attribute increase to include *sliced-shape*. So, the argument goes, the fact that we may not initially accept a sliced lemon as a lemon, may be attributable to an impoverished concept of *lemon*. The fact that we

later come to accept that lemons can be sliced, demonstrates the fact that this concept of *lemon* has changed.

Similar arguments could be raised in the other cases. The example of the tree struck by lightning shows that, initially, our prototype concept does not include *S-shaped* as a possible value of the attribute SHAPE. However, once we are made aware that *S*-shaped trees are trees nonetheless, our prototype concept of *tree* changes so as to include this possibility. That is, the change in judgement results from a change in the concept *tree*, not in the acquisition of the concept *struck by lightning.*

My response to this particular objection has already been hinted at and it runs as follows. No doubt there are concepts for which the values of attributes are specified disjunctively. The fact that apples can be red or green seems an obvious candidate. So, the argument is not one that suggests the values of attributes *cannot* be specified disjunctively. This is not a struggle against disjunction. Rather, the claim is that there are many cases where the values of attributes are *related.* The fact that apples may be brown (brown through being rotten, that is) is a case in point. Here the relationship between the possible colours of apples, red and brown, say, is qualitatively different to that between apples being either red or green. In the former case, there is a causal relationship to which we are attuned: we have words to describe the process, *over-ripe*, *bruised*, *rotten.* In the latter case there is none. In categorising a rotten, brown apple as an apple the claim is that we do not simply think of this individual as a brown apple even though we, perhaps, could: rather we think of it as a *rotten*, brown apple. To the degree that a theory of concepts purports to be concerned with thought, any such theory must reflect facts like these.

So, to repeat, though we could specify the fact that apples may be brown through being rotten by specifying *brown* as a possible COLOUR in the prototype concept *apple*, the argument is that a psychological theory of concepts should not. Categorisation, as Murphy & Medin have so ably argued, is sensitive to the causal relationships to which we are attuned. Any theory of concepts must respect this fact.

From the standpoint of development, prototype theory fares poorly in explaining how change in one concept may effect changes in other concepts. So, for example, it paints a picture of concept development in which information is accrued by static, isolated concepts. It does not indicate how the presence of one concept, *slicing* for example, may affect the applicability of others, *lemon* say. Such a failure, as our arguments suggest, must be seen as a significant inadequacy. We will suggest later (section 7.3) that one reason for this inadequacy is the fact that prototype theory is too abstract a statement of a theory of concepts.

7.2.3 Representational Economy

In this section we will mention two aspects of prototype concepts which appear to render them uneconomical in terms of representation. The first relates to the interaction among concepts that we have discussed in the previous two sections. The second relates to the significance of the way in which values are assigned weightings.

As we saw in section 7.1.1, a prototype concept must specify all the possible values which the attributes of that concept may assume. So, for instance, if it is possible that apples may be blue, as Putnam's arguments (Putnam, 1975) demonstrate, then the concept must specify *blue* as a value for the attribute COLOUR. But, of course, it is not simply the concept *apple* which must represent such possibilities. Presumably, our concepts for *lemon*, *tree*, *tiger*, etc., must also specify the possibility that at least one of their instances could be coloured blue. Indeed, it is difficult to place constraints on what we may call the *proliferation of possibilities* for concepts. What are all the possible values of all the relevant attributes for all concepts? Is the number of possibilities finite?

While such a question is difficult to answer, there are good reasons to suppose that prototype theory does not have the right one. For prototype theory would maintain that each prototype concept will specify the possibilities pertaining to its instances. For example, the prototype concept offered in Table 7.2, would presumably be further specified to include all possible colours, all possible shapes and all possible textures, not

to mention sizes, tastes, etc. Yet the puzzling fact is that most of these possibilities will also be specified by other concepts, concepts such as *car*, *orange*, *tiger*, etc.

Although it is possible that one way of avoiding this difficulty would be to invoke Cohen & Murphy's suggested mechanism of default inheritance for prototype concepts (perhaps, the value *flat* being inherited from the concept *object*), there seem good reasons even for not accepting this "solution". As has already been suggested in the two previous sections, the fact that, say, lemons may be flat is not best thought of as just one more possibility for the concept *lemon* to specify. Rather, this possibility is best expressed as the product of our concept *lemon* in conjunction with other concepts, say, that for *flattened*. Accordingly, such possibilities are not encoded in all concepts for which they are possibilities, nor in some concept superordinate to those. Rather, they are encoded in the concepts which express those very possibilities. They are encoded in concept such as *flatten*, *lightning-strike*, *slice*, etc.

This, then, shows one way in which prototype theory postulates uneconomical representations. Facts concerning the possible flattening of various objects should be encoded in concepts such as *flatten*, not in each and every concept some of whose instances may be flattened. So, prototype theory seems committed to there being a reasonable degree of redundancy of information in concepts. That is, if each prototype concept some of whose instances can be flattened, specifies this possibility, then essentially the same information, namely the possibility of flattening, is repeated. So, this is one sense in which prototype theory can be said to postulate uneconomical representations. The other sense in which this claim can be made, concerns the fact that prototype theory is committed to the view that values are weighted.

The weightings that are assigned to the values of a concept can be taken to reflect a number of things. Smith & Osherson, for example, suggest that, among others, such weights reflect the subjective probability associated with that value: the subjective probability that an instance possesses that value given that it is an instance of the concept. Cohen & Murphy assume that their ordinal ordering reflects a measure of family resemblance and that this measure is derived in the following way: sum the

number of features an instance has in common with other instances and subtract the number it shares with non-instances.

This view is clearly related to the notion of cue validity, roughly, the probability that given an entity has a certain feature, the entity is of a certain type. Let us consider an example: the value *four* for the attribute NUMBER-OF-LEGS with respect to the concept *dog*. Let us assume that the number of relevant instances and non-instances are a, b, c and d as follows.

	Dogs	Non-dogs
Four-legged	a	b
Non-four-legged	c	d

Now, the conditional probability that something is a dog given that it is four-legged, is equivalent to $\frac{a}{a+b}$. The conditional probability that something is four-legged given that it is a dog, is equivalent to $\frac{a}{a+c}$.

Now, the problem with views of prototype concepts that assume weightings on values reflect some notion of cue validity or subjective probability is simply that these *change*. They depend on the number of instances encountered and on the number of non-instances. So, for example, the probability that an instance is a dog given that it is four-legged depends not only on the number of four-legged dogs encountered but also on the number of four-legged non-dogs encountered. The task for a child in the process of acquiring a prototype concept is to maintain, for each and every such concept, the number of instances and non-instances encountered. Such a task, as Cohen (1983) notes, seems at first glance to be "near-impossible" (p. 87).

This, then, is the other sense in which prototype theory can be said to postulate uneconomical representations. By supposing that the ordering between the values of attributes is based on notions of cue validity or subjective probabilities, the only way in which such weights can change over time is if the numbers of instances and of non-instances of categories are recorded. This conclusion applies equally to Smith & Osherson's model in which weights may be free-valued and to Cohen & Murphy's model in which values are

effectively listed.

7.2.4 Central Exemplars

This section concerns a virtue of prototype theory, namely its treatment of the central tendency of categories. Prototype theory came into being as a result of the seminal work of Rosch. In a series of experiments she demonstrated that the central tendency of categories exhibited what we might think of as enhanced cognition. So, for example, subjects' judgement of whether or not various instances of bird fall under the concept *bird* are made reliably and significantly faster in the cases of typical birds as opposed to atypical birds. Robins, being fairly typical birds, are judged to be birds faster than, say, an ostrich. Indeed, on the basis of these reaction time studies, one can deduce an ordering among exemplars, from those eliciting the shortest response time to those eliciting the longest. In another experiment, Rosch demonstrated that subjects will quite readily order instances of a category according to their judged typicality. Though it is perhaps unsurprising, it has always been taken as strong support for prototype theory that these two, differently produced orderings of instances, often exhibit remarkable correlations.

The view that emerges from Rosch's work is that various aspects of our cognizance of atypical exemplars is dependent on our cognizance of the typical ones. For example, the time taken to judge whether an entity is an instance of a concept, depends on how similar the entity is to the typical exemplars of that concept. More properly, though, it depends on the similarity between the entity and the central tendency of the category, what has become known as the prototype. So, prototypes, themselves, are not necessarily descriptions of any particular exemplar, but they are descriptions embodying the most typical attributes and values associated with the category.

Rosch's work has assumed a central position in the psychology of concepts but it seems likely that in a very important sense, it has been misunderstood. Lakoff (1987) distinguishes between the prototype *effects* which Rosch's experiments demonstrate, and the

claims of representation made on their basis. He distinguishes two interpretations of prototype effects.

1. *The Effects = Structure Interpretation:* Goodness-of-example ratings are a direct reflection of category membership.

2. *The Prototype = Representation Interpretation:* Categories are represented in the mind in terms of prototypes (that is, best examples). Degrees of category membership for other entities are determined by their degree of similarity to the prototype.

Lakoff's points are, one: that most versions of prototype theory assume one or other, and sometimes both, interpretations of prototype effects. Two, that both such interpretations are not entailed by the existence of prototype effects.

Both points are valid. Prototype effects cannot be directly revealing of mental representation as interpretation 2 would have us believe. Similarly, the issues of goodness-of-example ratings and membership are logically distinct. The work of Armstrong, Gleitman & Gleitman (1983), for instance, demonstrates that subjects will readily order instances according to their perceived typicality, regardless of the facts of their category membership. In particular, one of Armstrong, Gleitman & Gleitman's tasks involved asking subjects for typicality judgements for odd numbers. So, they would be asked to rate various odd numbers for their typicality of the category of odd numbers. The fact that subjects do this, while knowing full well that the category has sharp boundaries, that there are clear conditions of membership and that membership itself is either all or none, indicates the fallacy in assuming that goodness-of-example ratings directly reflect graded membership.

Lakoff's intention is to argue against the proposal we have already mentioned, namely that of partitioning concepts into a core and a set of identification procedures. The motivation for such a partition comes from studies such as Osherson & Smith (1981) and Armstrong, Gleitman & Gleitman (1983) which purport to demonstrate that prototype

theory cannot be a complete theory of concepts. The rationale of these studies is to show that there is more to conceptual structure than that revealed by prototype effects. Correspondingly, the claim is that there is more to conceptual structure than prototype concepts. However, as Lakoff points out, this rationale is based on the two fallacious interpretations of prototype effects. So, Lakoff's argument is that versions of prototype theory like that of Smith & Osherson (1984) are not the only conceivable versions of prototype theory and, correspondingly, one cannot discount prototypes as playing a role in the central aspect of a theory of concepts, thought.

Our discussion in this section, then, has been to reflect on some of the original motivation for prototype theory, namely robust prototype effects. Some theorists have erroneously interpreted these effects as directly revealing of various aspects of conceptual structure and representation. Consequently, these theorists have been led to the erroneous view that it is not possible to characterise adequately a theory of concepts solely in terms of prototypes. However, these arguments show, instead, that it is not possible for these theorists to obtain such a characterisation in terms of their *models* of prototype effects. Such arguments do not count against the position that prototypes do play a central role in theories of concepts. They count only against existing models of prototype concepts.

Our conclusion, then, is that the original results of Rosch in demonstrating prototype effects, the crucial role of the central tendency in categorisation, is unscathed by the inadequacies of the models we considered earlier (section 7.1.1). Indeed, the claim still stands: any theory of concepts must account for the importance of the central exemplars, the central tendency of a category.

7.2.5 Contextual Sensitivity

The models of protoype theory detailed in section 7.1.1 share the property of representing what is known as graded structure. Graded structure refers to the fact that subjects are capable of discriminating members of a category on the basis of how typical they are of that category. Similarly, non-members of a category can also be ordered according

to how typical they are of that category. So, for example, with respect to the category of birds, subjects may generate the following order of most to least typical: sparrow, penguin, butterfly, golf ball. In Smith & Osherson's "Selective Modification Model", such graded structure is represented, in part, by the weights attached to the different values: *flies* would receive a higher weight than *does-not-fly* so as to discriminate robins from penguins; *has-beak* would discriminate penguins from non-birds such as butterflies. And so on. Similarly, in Cohen & Murphy's "Knowledge Representation Model", graded structure is reflected in the fact that values are ordered.

The observation of graded structure seems, then, to be part of the motivation of these models. Hence, it is all the more surprising that these same models cannot account for some notable findings regarding graded structure. They come from Barsalou (1987).

Barsalou summarises a number of findings which suggest that graded structure is not simply a function of similarity to the central tendency. Barsalou (1985), for instance, demonstrated that similarity to the central tendency failed to determine the graded structure of goal-derived categories such as *things to eat on a diet*, *things to pack in a suitcase*. Also ideals are implicated in determining the graded structure of categories. For example, the graded structure of *things to eat on a diet* is partly determined by how similar an instance is to the ideal thing to eat on a diet, namely something with zero calories. Barsalou also found that a major determinant of graded structure is how frequently an exemplar is perceived as instantiating its category. As perceived frequency of instantiation increases so does typicality. Despite the obvious interest of this for the models of prototype concepts that we have considered we will focus on a different determinant of graded structure, context.

Context, broadly construed, can influence the graded structure of categories in a number of ways. Firstly, in different contexts subjects seem to differentially employ the different determinants of graded structure we noted above. So, for example, in certain contexts subjects used ideals to determine graded structure, in other contexts, and for the same category, they would use the category's central tendency. Similarly, Roth & Shoben (1983) found that the graded structure of *animals* depends on linguistic context. In the

context of *milking* a cow is more typical than a horse. In the context of *riding*, this order is reversed. Perhaps one of the most interesting demonstrations of the instability of graded structure concerns its determination by subjects' point of view.

Barsalou & Sewell (1984) asked subjects to rate various instances of appropriate categories according to a point of view. These points of view were chosen according to nationality and according to occupation. For example, subjects were asked to rate various birds according to their typicality of the concept *bird*, and from, say, an American or a Chinese point of view. According to these different perspectives, different graded structures were indicated. For example, *eagle* is typical from an American perspective while atypical from a Chinese perspective.

The weight of evidence supporting the instability of graded structure, renders the models of section 7.1.1 impoverished, at least, as a theory of concepts. For, to capture such facts as we have considered, these models would have to specify some means by which the weights assigned to values may alter. In the absence of such a specification, then, such facts can only act to disconfirm the hypothesis that such models are models of concepts, even, indeed, of identification procedures. One can go further, in fact, and claim that the whole notion that weights based on subjective probabilities and cue validities can change with context is simply incoherent. For it is a property of such probabilities and cue validities that they do not change as a function of context. They should change only as a function of the number of instances and non-instances encountered.

The considerations of this section provide compelling evidence that any theory of concepts must explain the variability of the content that is ordinarily associated with the same concept on different occasions of use. This, then, is one of the standards by which the view of the next section must be judged. However, before we conclude this section on some of the deficits of prototype theory, let us briefly return to Barsalou for, in his description of the nature of concepts, we will find a number of interesting parallels with the view we present in the next section.

One important question for a theory of concepts is how they are to be mentally rep-

resented. According to Barsalou there are two obvious possibilities. The first is that concepts are represented as invariant structures in long-term memory and that these representations are accessed on those occasions that the concept is required. Consequently, the very same conceptual content should be retrieved on all such occasions. The second possibility is that concepts are represented as constructions in working memory, the information from which they are constructed being drawn from (possibly) invariant representations in long-term memory.

The first possibility reflects most traditional views of concepts. Concepts are seen as having a fixed content and this content is accessed whenever the concept is needed. Such a view, however, has marked difficulty in accounting for the facts concerning contextual variability that we have noted, namely, that people tend to associate different information with the same concept on different occasions of use. Indeed, we have suggested in Chapters 5 and 6 that, with respect to word meaning, the problem of accounting for the variability of content is created by adopting a traditional analytic view. It is the very nature of this traditional analytic conception of meaning that meanings are not sensitive to context, that one and the same meaning holds globally. And, indeed, a similar conclusion can be applied to theories of concepts.

At any rate, given this problem, Barsalou assumes that the second possibility for a theory of the source of concepts is the correct one. Thus, according to his view, conceptual content arises from a constructive process which yields representations in working memory from those stored in long-term memory. However, Barsalou, perhaps unsurprisingly, regards the traditional analytic view of concepts as genuinely mistaken. Whatever else such a a view may be a view of, it is not to be mistaken for a view of concepts. This is what he says.

> Consequently, the concepts that theorists "discover" for categories may never be identical to an actual concept that someone uses. Instead, they may be analytic fictions that are central tendencies or idealisations of actual concepts. Although such theoretical abstractions may be useful or sufficient for certain scientific purposes, it may be more fruitful and accurate to describe the variety of concepts that can be constructed for a category and to understand the process that generates them.

> It should be noted that this argument is *not* meant to imply that there is *no* stable knowledge in long-term memory. Instead, the point is that the same exact representation does not appear to represent a category on every occasion in working memory. As discussed in the following section, it is entirely possible for there to be relatively stable knowledge in long-term memory, but for it rarely to be the case that the same information is retrieved from this knowledge to represent a category.
>
> *(Barsalou, 1987; p. 120)*

Barsalou's claim, then, is that concepts, as traditionally conceived, are "fictions". Though there may be some knowledge stably represented in long-term memory, these are almost certainly not to be thought of as representing conceptual content. Rather, the content that we associate with concepts is generated from such stable representations according to some mechanism which is sensitive to, among others, context. In Chapter 4 we presented a view of senses which was motivated, in part, by observations such as Barsalou's and by the suggestion that senses are mentally represented in terms of concepts. Hence, that view can be seen, quite naturally, as a view of concepts. In the next section we briefly re-iterate this view and enumerate the ways in which it purports to avoid the pitfalls of prototype theory. Having done this, we will attempt a detailed comparison with prototype theory.

7.3 The Family of Constraints View

The Relational View that we developed in Chapter 4, the view which we now call the Family of Constraints or FoC view, makes a number of commitments. One is to the view that underlying the various uses of, what is traditionally thought of as, a single concept are different combinations of concepts. All of these involve combinations with a concept which corresponds to the central tendency of the category traditionally associated with the single concept. Thus, the central concept may encode information pertaining to the prototype but, importantly, it will not be a prototype concept in the sense intended by prototype theorists. Let us consider an example, the concept *lemon*. Just as we did in Chapter 4, we will describe the example both in terms of feature structures and in terms of the conditional constraints of Situation Theory. And, just as before, the claims of the

$$\begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{oval} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix}$$

Table 7.3: A Possibility for the Contents of the Concept *lemon*.

theory must be seen as less to do with the precise contents we associate with concepts, and more to do with the mechanism by which those contents may change. Table 7.3, then, details one possibility for the concept *lemon*.

Now, we can express such facts in terms of Situation Theory, firstly by assuming that there is some relation between the type of uses of the word *lemon* and some properties, P, given by the concept *lemon* and, secondly, by assuming that this conceptual content is related to conditions under which the word *lemon* may be used. The first assumption is expressed as follows. Just as before this relation is a WORM.

$$\ll \text{WORM, LEMON, P; } 1 \gg$$

where

$$\begin{aligned} \text{P} = \quad [\dot{\text{p}} \mid \quad & \ll \text{fruit, } \dot{\text{p}}\text{; } 1 \gg \\ & \ll \text{yellow, } \dot{\text{p}}\text{; } 1 \gg \wedge \\ & \ll \text{oval, } \dot{\text{p}}\text{; } 1 \gg \wedge \\ & \ll \text{acidic, } \dot{\text{p}}\text{; } 1 \gg \wedge \end{aligned}$$

The second assumption is expressed by the following conditional constraint.

$$S_1 \Rightarrow S_2 \mid B$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll \text{P, } \dot{\text{p}}\text{; } 1 \gg]$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \textit{lemon}; 1 \gg]$$

and

$$B = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LEMON}, P; 1 \gg]$$

and

$$\begin{aligned} P = [\dot{p}_1 \mid\ & \ll \text{fruit}, \dot{p}_1; 1 \gg] \\ & \ll \text{yellow}, \dot{p}_1; 1 \gg \wedge \\ & \ll \text{oval}, \dot{p}_1; 1 \gg \wedge \\ & \ll \text{acidic}, \dot{p}_1; 1 \gg \wedge \end{aligned}$$

Now, in many respects, the conceptual content illustrated in Table 7.3 seems to exemplify the classical theory of concepts. But this is a deception, for, as we saw in Chapter 6, the classical theory of concepts is committed to the traditional conception of analyticity. Namely, that concepts should express necessary and sufficient conditions for their application. This view, however, is committed to the notion that a concept's content may express conditions which are only contingently true of that concept's application. This forces us to be extremely careful in talking of what it means for an entity to fall under a concept.

All manner of weird and wonderful lemons will not satisfy the description given in Table 7.3, yet, it is clear, that we need to say that such entities nonetheless fall under the concept of *lemon*. In many respects, this tension is reminiscent of the discussion of 1.3 where we distinguished two claims of concepts: a functional claim, and a representational claim. There, we argued that a functional classification of behaviour, such as concepts are deemed to effect, cannot lead directly to a theory of how such classifications are to be mentally represented. Indeed, ignoring this fact can lead to theories of concepts

$$\left(\begin{bmatrix} \text{SHAPE:} & X \\ \dots & \dots \end{bmatrix}, \begin{bmatrix} \text{SHAPE:} & \textit{flat} \\ \dots & \dots \end{bmatrix}\right)$$

Table 7.4: A Possibility for the Contents of the Concept *flatten*. The ...indicate attributes and values which are repeated or inherited across the two arguments to the concept.

falling into error. Well, the question of what it means to fall under a concept requires a similar answer.

If, as we hypothesise, the contents of the concept *lemon are* mentally represented, even though we may not specify *how* they are represented, then falling under the concept can be determined according to how you interpret this theory of concepts. If, for example, you take concepts to be those stable structures which a theory hypothesises to be represented in long-term memory, then what it means to fall under the concept *lemon* is simply to satisfy the contents exemplified in Table 7.3. That is, if an entity falls under the concept *lemon*, that entity must be a yellow, acidic-tasting, oval-shaped fruit. However, such a view of what a concept is seems far too narrow for our purposes. And falling under a concept is similarly more complex than suggested.

As Barsalou has indicated, concepts are associated with different contents on different occasions of use. This is precisely the view which emerged in Chapter 4 and, consequently, we shall say that what it means for an entity to fall under a concept is for that entity to satisfy the content associated with the concept on a particular occasion of use. In order to examine this in a little more detail, let us consider how it is that the concept *lemon* can come to be associated with different content.

As Murphy & Medin's arguments suggest, our knowledge of operations and transformations are likely to be implicated in the content of concepts. Indeed, such a view can be seen as giving credence to the view presented in Chapter 4. There, we considered a number of relations, some of which correspond to such operations. One was the concept *flatten*, whose contents are exemplified in Table 7.4.

Again, the concept *flatten* can be described in situation-theoretic terms as follows. We assume that there is a WORM relation holding between the type of uses of the word *flatten* and two properties given by the concept *flatten* One property will apply to individuals of indeterminate shape, the other to individuals with exactly the same properties *except* that of having flat shape. Now, when the concepts for *lemon* and *flatten* combine, that is, by the concept *lemon* standing in the first argument of the concept *flatten*, the second argument of the concept *flatten* assumes the following conceptual content.

$$\begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{flat} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix}$$

Such a content also expresses attunement to a situation theoretic conditional constraint as follows.

$$S_1 \Rightarrow S_2 \mid B'$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg]$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \textit{lemon}; 1 \gg]$$

and

$$B' = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LEMON}, P_1; 1 \gg \wedge \ll \text{WORM}, \text{FLATTEN}, P_1, P_2; 1 \gg]$$

and

$$\begin{aligned} P_2 = [\dot{p} \mid & \ll \text{fruit}, \dot{p}; 1 \gg] \\ & \ll \text{yellow}, \dot{p}; 1 \gg \wedge \\ & \ll \text{flat}, \dot{p}; 1 \gg \wedge \\ & \ll \text{acidic}, \dot{p}; 1 \gg \wedge \end{aligned}$$

So, as a consequence of the process of combining of concepts, a new conceptual content, one that is not hypothesised to be stored in long-term memory, is generated. And this content expresses attunement to a different though related constraint. This gives us a hint as to the more sophisticated answer as to what it means for an entity to fall under a concept. We will tentatively suggest that an entity can be said to fall under a concept X if that entity satisfies the conceptual content associated with concept X on an occasion of its use. So, in the case of an entity falling under the concept *lemon* it is not simply a matter of that entity falling under the contents hypothesised to be represented in long-term memory. Rather, it is constituted by that entity falling under either the contents of the mentally represented concept or the contents generated from that concept by some process of concept combination. This, then, allows us to say that flattened lemons fall under the concept *lemon* even though they do not satisfy the content of the concept *lemon* which is deemed to be represented in long-term memory. Whether such a view of what it is to fall under a concept can be sustained is, I assume, an empirical matter and one worthy of further investigation. Needless to say, this is a matter for further research and not for the present goal of articulating a view of word meaning and concepts.

However, there are one or two obvious objections to this view which need to be answered. Firstly, it has been suggested that there all manner of possible concept combinations, some of which apply to entities which cannot be said to fall under a constituent concept. For example, if we are committed to the view that flattened lemons fall under the concept *lemon*, surely we are as committed to the view that fake guns fall under the concept *gun*. This, it is argued, is absurd. On the contrary, I think this is exactly what a theory of concepts should strive to attain.

It must be remembered that under the view I am proposing, falling under a concept is a contextually sensitive process. Only in certain situations will the constraints to which concepts express attunement actually hold. So, in certain situations, if one were to claim that *fake gun* falls under the concept *gun*, one would be wrong. The claim would be misinformative. However, in certain other situations, the claim that *fake gun* falls under the concept *gun* would, in some way, be correct: it would be informative. That is, to claim that a fake gun is a gun would not, in *all* circumstances, simply be a false assertion. In many cases, it can carry information. For example, consider a situation where there are two guns in the room, one real, one fake. And suppose that one needs a gun for a certain purpose such that it does not matter which of the two guns you choose. Making a drawing, for example. Or, provided the real gun is unloaded, for using as a prop in a play. Under these circumstances, to ask for a gun is not to ask for a real gun. Nor, if one was handed a fake gun, would the claim that one had not been handed a gun be felicitous.

Similarly, in the Lion Puzzle of Chapter 1, Fred assents to the claim that a stone lion is a lion. In the same sense in which fake guns are taken not to be guns, stone lions are also not lions. But the fact is that the word *lion* can, under certain conditions, apply to stone lions, just as *gun* can apply to fake guns. Indeed, we have suggested that the word *lion* may have a sense "stone statue of a lion", and we are similarly committed to the view that the word *gun* may have a sense *fake gun*. As we noted in section 5.1.1, Evans (1976) has argued that the intuition behind the Fregean conception of sense is that a sense is a way of thinking of the entity or entities to which that sense applies. Accepting this position, then, commits us to the notion that there are numerous different ways of thinking of an object, these ways being associated with the same linguistic string. So, for example, the uses of the word *lion* betray numerous ways of thinking of the objects to which that word refers. And this returns us to our discussion of concepts.

By assuming that there are numerous different ways of thinking of one and the same object or group of objects, we are committed to the notion that associated with what is traditionally thought of as a single concept, are numerous different conceptual contents. So, for instance, we can think of a flattened lemon as a lemon because, under certain

circumstances, the concept *lemon* is associated with a content satisfied by flattened lemons. However, it may well be the case that, in certain other circumstances, the concept *lemon* is not, that is not in those circumstances, associated with such a content. In those very same circumstances, then, the view I am proposing is committed to the view that it is possible to think of flattened lemons as not falling under the concept *lemon*. Whereas some claim that this is absurd, I would disagree. In the case of fake guns, I would claim that it is perfectly acceptable to claim that we can view these both as guns and not as guns, but that how we view them on any particular occasion depends on the circumstances of that occasion.

So, the retort to this particular objection is that it *is* possible, under certain circumstances, even in the cases of privative combinations such as *fake gun* and *stone lion*, to assert that entities of this type are also of the head noun type. That is, they are also guns and lions, respectively.

Another related objection is that while squashed lemons undoubtedly are real lemons, stone lions are not real lions and that this difference should be accorded a distinction in their theoretical treatment. Firstly, our theory does accord this difference a theoretical distinction. It is assumed that only certain contexts facilitate a use of *lion* to refer to real lions, and only certain other contexts facilitate a use of *lemon* to refer to flattened lemons. Though we have not been precise about the nature of these differences in context, our position is quite consistent with results that suggest that there *is* some intrinsic difference between these two sorts of context. Secondly, as we argued in section 1.1, it is important to distinguish the conditions under which an entity can be *said* to be an X and the conditions under which an entity *is* an X. This second objection presumes that these two questions will yield the same answer. So the retort to this objection is that we do not make such a presumption.

We are now in a position to enumerate some of the ways in which this view avoids the problems that have seemingly beset prototype theory.

1. **Coherence:** Conceptual coherence is determined, in part, by explanations of the

sort Murphy & Medin describe. The fact that the concept *lemon* can only apply to flattened lemons if there is some concept which furnishes an "explanation" as to why flattened lemons are not oval as the concept *lemon* indicates.

2. **Developmental Adequacy:** The application of concepts is determined, in part, by the acquisition of other concepts. Not having a concept such as *flatten* precludes the application of the concept *lemon* to flattened lemons.

3. **Representational Economy:** Proliferation of possibilities is avoided by allowing such possibilities as flat shape for the concept *lemon* to arise from the *combination* of concepts. Possibilities are not encoded in each concept for which they *are* possibilities as they are in prototype theory. The resource overheads associated with weighted values are avoided by reverting to an apparently classical means of conceptual representation.

4. **Central Exemplars:** The central exemplars, or central tendency of a category, are accorded a special cognitive significance. Indeed, concepts may well represent prototypes, though not prototype concepts. It is left to empirical studies to decide whether a particular concept should represent prototypes, ideals, paragons, etc.

5. **Contextual Sensitivity:** The fact that numerous different conceptual contents are associated with what is traditionally thought of as the same concept, offers an explanation for the fact that the contents of concepts seem highly context sensitive.

One query which must be raised for a view of concepts such as this concerns the absence of weighted values. Within prototype theory such weights are taken to reflect, and sometimes taken to explain, the prototype effects we discussed earlier. Precisely how our view explains such prototype effects we will see in the course of the next section where our focus is on comparing the view we have described directly with prototype theory.

7.3.1 A Comparison with Prototype Theory

A comparison between prototype theory and the FoC view might seem, at first glance, problematic. Firstly, according to prototype theory the application of a concept is determined by one constraint, according to our view it is determined by many such constraints, indeed a whole family of related constraints. Secondly, protoype theory seems committed to the view that concepts express attunement to unconditional constraints. Our view is that these constraints are conditional. However, we will attempt a comparison, nonetheless. The means by which we approach this task is by way of attempting to modify the view we have developed in order that it may express the sorts of prototype concepts modelled by Smith & Osherson and Cohen & Murphy. The nature of the modifications we will have to make, will determine how we are to interpret prototype theory.

Firstly, let us introduce a number of abbreviations so that we may describe our view in as economical a way as possible. Let us abbreviate the situation type where x has the property P to PROP(P, x) and the situation type where x can be described by a use of the word *lemon* or, alternatively, where x can be said to fall under what would traditionally be thought of as a concept, to DESCRIBE(x, *lemon*). Similarly, WORM(LEMON, P), B_1, etc., all refer to situation types. What should be clear, at any rate, is that this notation is not intended to convey formal precision but, rather, to simplify what we do intend to convey which are the bones of the frameworks under consideration.

Now, the view we have developed is committed to the view that underlying the various uses of a word and the domain of application of what are traditionally though of as concepts, are multiple conditional constraints. In the case of prototypical lemons, lemons which have been flattened and lemons which have been painted blue, these are, respectively, as follows.

PROP(P, x) ⇒ DESCRIBE(x, *lemon*) | WORM(LEMON, P) ∧ B_1

PROP(Q, x) ⇒ DESCRIBE(x, *lemon*) | WORM(LEMON, P)∧
WORM(FLATTEN, P, Q) ∧ B_2

PROP(R, x) ⇒ DESCRIBE(x, *lemon*) | WORM(LEMON, P)∧
WORM(PAINT, P, R) ∧ B_3

Now, our first task in attempting to reconstruct prototype theory is to attempt to convert these conditional constraints into unconditional constraints. For conditional constraints cannot be thought of as complex "conjunctive" constraints unless the background conditions with respect to which each are relativised are identical. Hence, we must first try to establish unconditional constraints equivalent to those above.

One way of making conditional constraints unconditional is to do what Baker (1974) calls "conditionalisation". This involves reducing a constraint such as

$$S_1 \Rightarrow S_2 \mid B$$

to the "logically" equivalent form

$$S_1 \wedge B \Rightarrow S_2$$

Now, there are good reasons to avoid such a reduction. As Baker notes, Wittgenstein explicitly denies the efficacy of such a move. Barwise (1989) suggests that this reduction is not valid, although he suggests that there is a reduction of a more complex nature. In Braisby & Franks (in preparation), it is argued that this "conditionalisation" mistakes the nature of conditional relations. In particular, it is argued that certain defeasible relations may be considered *reducibly* defeasible. That is, they may be recast as conditional relations. Other defeasible relations, however, are *irreducibly* defeasible and cannot be reduced to conditional relations: they remain intrinsically defeasible and are inherently unreliable. Conditionalisation casts *all* conditional relations as defeasible ones, hence

losing the important distinction between conditional or reducibly defeasible relations and irreducibly defeasible relations.

So, there are good reasons against this "conditionalisation" but, if we are to reconstruct prototype theory from the view we have developed, then this is a move we must make. Hence, at this stage, intermediate between prototype theory and our own view, the picture we have is as follows.

$$\begin{aligned}
\text{PROP}(P, x) \wedge \text{WORM}(\text{LEMON}, P) \wedge B_1 &\Rightarrow \text{DESCRIBE}(x, \textit{lemon}) \\
\text{PROP}(Q, x) \wedge \text{WORM}(\text{LEMON}, P) \wedge & \\
\quad \text{WORM}(\text{FLATTEN}, P, Q) \wedge B_2 &\Rightarrow \text{DESCRIBE}(x, \textit{lemon}) \\
\text{PROP}(R, x) \wedge \text{WORM}(\text{LEMON}, P) \wedge & \\
\quad \text{WORM}(\text{PAINT}, P, R) \wedge B_3 &\Rightarrow \text{DESCRIBE}(x, \textit{lemon})
\end{aligned}$$

Now, our avowed goal is to reduce these three constraints to a single constraint to which a prototype concept might express attunement. If we re-examine the models of prototype concepts we discussed earlier, it is clear that, in these models, no mention is made of other *concepts* like those explicitly shown in the above constraints. No reference is made, for example, to the concept *flatten*, though the value *flat* is represented and, indeed, this may even represent the concept *flat*. However, this value *flat* is encoded in terms of the property Q. The additional concepts in the constraints indicated above merely serve to indicate the "origin" of such values and, as we saw in section 7.2.1, this is not something that prototype concepts encode. Consequently, one task we must perform is to eliminate this dependence on such coherence-inducing concepts. If we eliminate such concepts from our constraints then we will come closer to specifying a relationship simply between the fact of a concept applying, and the properties an object must have for that concept to apply. That is, we will have constraints such as the following.

$$\begin{aligned}
&\text{PROP}(P, x) \wedge B_1 \Rightarrow \text{DESCRIBE}(x, \textit{lemon}) \\
&\text{PROP}(Q, x) \wedge B_2 \Rightarrow \text{DESCRIBE}(x, \textit{lemon}) \\
&\text{PROP}(R, x) \wedge B_3 \Rightarrow \text{DESCRIBE}(x, \textit{lemon})
\end{aligned}$$

Now we have three simpler constraints, ones relating properties to facts concerning the

application or otherwise of a given word. That is, corresponding to the three categories, normal lemons, flattened lemons and painted lemons, we have three unconditional constraints. Notice, that in making this last move, we have lost all information concerning why these properties and not others are implicated in determining the conditions of application of the concept. That is, we are now at a loss to explain the important role that theories play in the psychology of categorisation. But we are still far from our goal of obtaining a constraint to which a prototype concept might express attunement. In particular, there are two outstanding differences to be resolved. One concerns the fact that prototype concepts are unitary structures, they express attunement to single constraints, not the several that we have above. The other concerns the background situation types, B_1 to B_3. Let us consider these first.

The question we have to answer is whether a constraint of the form

$$P \wedge Q \Rightarrow R$$

can be reduced to the form

$$P \Rightarrow R$$

In some respects this is the question we should have answered in determining whether we could dispense with concept expressions from the previous set of constraints. Our answer then was in the affirmative simply because prototype theory is committed to the fact that a psychological characterisation of concepts can abstract away the effects of other concepts. So, if we are to reconstruct prototype theory then we must adopt such a move. However, it is distinctly unclear that prototype theorists would accept the reduction indicated above in general, or, even in the case of background situation types in particular. In fact, such a reduction will only be valid if Q is a necessary existent. That is, Q is true in *all* situations. So, in order to reduce our three constraints in a similar manner, we must be committed to the view that B_1 to B_3 are true of all situations. That is, every situation is of these three types. Now, such a view is almost

certainly false. One of the intuitions behind the view that we have developed was that the uses of concepts are extremely sensitive to context. In certain situations the concept is used with such-and-such content, in other, different situations it is used with different contents. That is, the approach assumes that B_1, B_2 and B_3 do distinguish different types of situation: there are situations which are not of each type.

To assume that all situations are of these types is to reject the notion that the contents of concepts are determined by context. One is left only with the notion that the content of a concept subsumes all the contents associated with its uses and, moreover, that the contents associated with those uses may not be exactly the content of the concept. Such a view seems tied to the idea that on all occasions of use the same content is accessed, even though it may not all be employed. However, this, too, is a move we are required to make. It results in the following set of constraints.

$$\text{PROP}(\text{P}, \text{x}) \Rightarrow \text{DESCRIBE}(\text{x}, \textit{lemon})$$
$$\text{PROP}(\text{Q}, \text{x}) \Rightarrow \text{DESCRIBE}(\text{x}, \textit{lemon})$$
$$\text{PROP}(\text{R}, \text{x}) \Rightarrow \text{DESCRIBE}(\text{x}, \textit{lemon})$$

In order to obtain a better approximation to prototype theory, we must attempt to reduce these three different constraints to one. Now, the supposition that there are constraints of a certain type, carries, at least, two commitments. Firstly, there is a logical commitment. That is, one must be committed to the view that the relata of the constraint are logically related in the way the constraint suggests. So, for example, suppose that the relation between being gold and having the atomic number 79 is best expressed by a necessary unconditional constraint. The logical commitment is, then, that the relationship between gold and this fact about atomic structure is one that holds in all circumstances and holds necessarily. If we are of the view that the relationship does not have these properties, then we will not view the relation in terms of a necessary, unconditional constraint. However, positing constraints carries a further commitment of an ontological nature. To suggest that there is an unconditional constraint relating the fact of being gold to the fact of having atomic number 79 is not just to specify the logical nature of the relationship between these facts, it is also to commit oneself to the view that situations contain not just these facts but also a fact that relates them, this

being a constraint. To see how it could be the case that just by considering the logical nature of constraints, we can make an unfortunate ontological commitment, consider the following example. Suppose, we consider two elements, gold and hydrogen. Now, gold has the atomic number 79 and hydrogen the atomic number 1. We might assume that there are two unconditional constraints relating each element to its atomic number. This, then, carries two commitments: the logical one, and the ontological one. Now, were we to combine these we would obtain a single constraint relating x being gold and y being hydrogen to x having atomic number 79 and y having atomic number 1. But notice that, while the logical characterisation of this combined constraint is no different in this case from the case where we have two separate constraints, an entirely different ontological statement is being made. Ontological grounds for accepting the ontological commitments made by each separate constraint will surely not be grounds for accepting the ontological commitments made by the combined constraint. For, in the latter case, our ontology must group together gold and hydrogen, and plausible reasons for such a grouping must be given. In the case of the independent constraints no such reasons are required. And, whereas plausible reasons *may* be available in the case of gold and hydrogen, we can render this unlikely by considering cases where ontological boundaries are crossed. For example, we could combine the constraint relating gold to its atomic number and a constraint relating elephants to some hypothesised elephant-essence.

Combining the above constraints, then, as we must do if we are to obtain a view resembling that of prototype theory, is likely to represent a not inconsiderable ontological pitfall. However, let us proceed in our attempt to reproduce a constraint to which a prototype concept might express attunement. The three constraints we have indicate that, associated with a concept, are three different contents. Prototype concepts, however, specify a single content which is associated with any particular use of a concept. The only way we have of combining these constraints is, therefore, via disjunction. That is, we will obtain a constraint that will tell us that either of three sets of conditions must hold for the concept to apply. The result we obtain, then, is as follows.

$$\text{PROP}(\text{P}, \text{x}) \vee \text{PROP}(\text{Q}, \text{x}) \vee \text{PROP}(\text{R}, \text{x}) \Rightarrow \text{DESCRIBE}(\text{x}, \textit{lemon})$$

We have now arrived at a picture of conceptual content which is expressible in terms of one, unconditional constraint. However, we have not yet arrived at the kind of constraint to which a prototype concept might express attunement. To see this we need to remind ourselves of the particular properties that P, Q and R abbreviate.

$$P = \begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{oval} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix} \quad Q = \begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{yellow} \\ \text{SHAPE:} & \textit{flat} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix}$$

$$R = \begin{bmatrix} \text{ISA:} & \textit{fruit} \\ \text{COLOUR:} & \textit{blue} \\ \text{SHAPE:} & \textit{oval} \\ \text{TASTE:} & \textit{acidic} \end{bmatrix}$$

Now, what our constraint suggests is a disjunction of complex conjunctive properties. And, as such, the constraint expresses what we might think of as "feature co-ocurrence restrictions". That is, suppose that one of our complex conjunctive properties, call it S, was satisfied by rotten lemons. Further suppose it is a fact that all normal rotten lemons not only change to a brown colour but also become sweet flavoured. Further suppose that the only way a lemon can be sweet flavoured is by being rotten. Under such conditions, a sweet lemon is always a brown one. That is, the presence of one feature places a restriction on the presence or absence of another. Now, it is a point made by Murphy & Medin (1985) that the kinds of model exemplified by those we discussed earlier do not express such facts as these feature co-occurrence restrictions. These, too, must be lost if we are to obtain something akin to a prototype concept.

Prototype theory specifies for each concept all the possible values that each attribute may assume. So, for example, whereas our constraint may specify the following

$$(acidic(x) \wedge yellow(x)) \vee (sweet(x) \wedge brown(x))$$

prototype theory seems to prefer the following form.

$$(acidic(x) \vee sweet(x)) \wedge (yellow(x) \vee brown(x))$$

As we have noted, these forms are truth-conditionally distinct. However, we now recast our above constraint into these terms. The result follows.

$$\begin{aligned}(\textit{fruit}(x) \vee \textit{fruit}(x) \vee \textit{fruit}(x)) \quad &\wedge \\ (\textit{yellow}(x) \vee \textit{yellow}(x) \vee \textit{blue}(x)) \quad &\wedge \\ (\textit{oval}(x) \vee \textit{flat}(x) \vee \textit{oval}(x)) \quad &\wedge \\ (\textit{acidic}(x) \vee \textit{acidic}(x) \vee \textit{acidic}(x)) \quad &\Rightarrow \text{DESCRIBE}(x, \textit{lemon})\end{aligned}$$

We have deliberately written this out in long-hand, redundant form, in order to point to the last outstanding difference between what we now have and the kind of constraint to which prototype concepts express attunement. Namely, the weights on values. We can clearly see that certain values are repeated and, while this makes no truth-conditional difference, it does suggest that certain attributes are more typical of the category. Thus, the final move of eliminating the truth-conditional redundancy, but introducing a notion of typicality based on the frequency of occurrence of attributes' values, yields the following.

$$\begin{aligned}\textit{fruit}\ (x){:}3 \quad &\wedge \\ (\textit{yellow}\ (x){:}2 \vee \textit{blue}\ (x){:}1) \quad &\wedge \\ (\textit{oval}\ (x){:}2 \vee \textit{flat}\ (x){:}1) \quad &\wedge \\ \textit{acidic}\ (x){:}3 \quad &\Rightarrow \text{DESCRIBE}(x, \textit{lemon})\end{aligned}$$

This constraint, then, is of precisely the kind suggested by the models of prototype concepts we discussed earlier. Notice, however, that this last move only makes sense if one believes that a concept should express abstractions across different contexts, even contexts in which the content associated with that concept is palpably distinct.

At any rate, it can be said that prototype theory can be reconstructed out of the FoC view we have developed. The picture which emerges, however, is that prototype concepts express abstractions over the uses of concepts to a degree that the content associated with uses of the concept are some subpart of the specification of the prototype concept. Presumably, prototype theorists must invoke some additional mechanism to explain why only a subpart of a concept's content appears to be employed on various occasions of use. It is to the benefit of the alternative view we have developed that a natural account

of this phenomenon is forthcoming. Before we finish this section, let us enumerate the difficulties encountered in reconstructing prototype theory. Although, the existence of these difficulties does not invalidate prototype theory, (one will, after all, only encounter them if one first adopts a view such as our own), they do suggest that our account may offer a more natural account of concepts. Here, then, are what we assume the difficulties we encountered suggest for prototype theory.

1. **Conditionalisation:** Prototye theory appears to rely on conditionalisation. That is, it relies on treating relations which are intrinsically conditional or reducibly defeasible as irreducibly defeasible. Prototype theory, then, seems committed to the view that concepts only express attunement to relations which are not conditional.

2. **Coherence:** Prototype theory appears to require abstraction over concepts and theories of which those concepts are a part. That is, it assumes that information concerning why it is that possible values are possible is dispensable as regards categorisation.

3. **Background Conditions:** Prototype theory requires us to assume that the application of a concept is essentially non-conditional. That is, there are no background conditions on which that application depends. Hence, we can dispense with the hitherto unknown situation types B_1 to B_3.

4. **Ontology:** Prototype theory risks an ontological pitfall by assuming that differing information, possibly relating to different ontological categories, may be conveyed by one and the same relation.

5. **Feature Co-occurrence:** Prototype theory seems committed to the view that concepts do not express feature co-occurrence restrictions. As such it is committed to there being *loss* of information pertaining to exemplars.

6. **Weighted Values:** Prototype theory seems committed to the notion that weights on values represent the frequency of occurrence of those values across all contexts. That is, prototype concepts represent context-independent aggregations of contextually sensitive information.

All of the above, then, are, from the perspective of the view we have developed, violations. We now turn away from our direct comparison with prototype theory and towards one of the much vaunted virtues of prototype theory, its account of family resemblance.

7.3.2 On Family Resemblance

Prototype theory has a fairly natural and easily understood account of family resemblance. Associated with each prototype concept is a cluster of descriptions or attribute-value pairs, not all of which need be true of every exemplar. For an exemplar to be an instance of the concept it need only receive a sufficient number of "votes". There is no commitment to the idea that every exemplar possesses some set of core or common features. In this, then, it is said that prototype theory accounts for family resemblance.

While this much is beyond dispute, I would like to suggest that an alternative account of family resemblance is forthcoming from the view we have developed. According to FoC, underlying the various uses of a word there are a number of related constraints. For example, underlying uses of the word *lemon* are constraints corresponding to normal lemons, flattened lemons, painted lemons, toy lemons, wax lemons, even stone statues of lemons, if circumstances require. A number of things are clear from such an account. First, that underlying all uses there need be no common properties to all those entities to which the word may apply. Second, that all these uses, all these constraints, are related.

The content associated with uses of *lemon* in referring to flattened lemons is *related*, via the concept *flatten*, to the content associated with uses of *lemon* in referring to normal lemons. The content associated with uses of *lion* in referring to stone lions is *related*, via the concept *statue*, to the content associated with uses of *lion* in referring to normal lions. The content associated with uses of *mother* in referring to adoptive mothers is *related*, via the concept *adoptive*, to the content associated with uses of *mother* in referring to biological mothers. And so on. This, I take it, is what Wittgenstein means in the quote with which we begin Chapter 4.

7.4 Conclusion

The conclusion of this discussion concerns the efficacy of prototype theory. We have argued that prototype concepts must be seen as expressing attunement to single unconditional constraints. We have presented a number of problems with standard models of prototype concepts. In contrast, we re-iterated the view we presented in Chapter 4 and we detailed reasons why such a view avoids many of the difficulties which face prototype theory. Having done so, we attempted a detailed comparison with prototype theory, the purpose of which was to suggest some underlying reasons for prototype theory's difficulties, perhaps the most important of which, is its inadequate treatment of context. Finally, we have pointed out that the view we have developed, the Family of Constraints view, offers an account of family resemblance which seems to be in keeping with Wittgenstein's observations.

I want to conclude this chapter, by returning to Barsalou. He claimed that concepts are nothing more than "analytic fictions". We are now in a position to say the following. According to the Family of Constraints view, prototype concepts are indeed fictions. They represent abstractions of content over all contexts of use. Concepts, as Barsalou noted, are best thought of as having content for an occasion of use. However, such an account should not be accused of lacking formality or precision. We have offered a formal interpretation of this account and, indeed, this framework has been cast into computational form (Braisby, 1990).

Barsalou has also claimed, some may think paradoxically, that the suggestion that what are traditionally thought of as concepts are not stored in long-term memory, does not imply that the content of concepts are not constructed out of invariant structures in long-term memory. Indeed, we have shown that this is not paradoxical for, while it is possible for what we call concepts to inhabit long-term memory and remain invariant, we have demonstrated how the content associated with their uses vary considerably with context.

Chapter 8

The Conclusion

In this chapter we conclude by casting a retrospective glance over the ideas presented in chapters 1–7 and by offering some pointers at future research which may further develop those ideas and tease out some of their implications. We begin by recapping. We end by pointing out some implications for future research.

This thesis has attempted to develop an account of word meaning which not only respects psychological arguments but also philosophical considerations. Part of the approach has been to express this account in terms of a formal semantics, that provided by Situation Theory, and to describe the informational contents associated with word uses in terms of feature-structures. It is hoped, then, that the thesis marks the beginning of a truly interdisciplinary, cognitive science of word meaning.

The thesis not only contains an argument, albeit a well-disguised one, but also much which cannot be argued for or, at least, much for which I cannot argue. There are various prejudices, inclinations, suspicions, hopes and hunches most of which concern the role that psychology should play in a cognitive-scientific study of word meaning. In the following, I hope to mark the boundary between my arguments and these other important aspects of scientific research.

8.1 The Arguments

In the first chapter we sketched the background to the rest of the thesis. Firstly, we attempted to do no more than hint at the location of this thesis in terms of a wider goal of characterising what Vygotsky calls verbal thought. Thus, we too are led to consider the meanings of words. That our enterprise concerns not only the behaviour of words but also the thoughts associated with their uses, has been made more explicit in later chapters.

One of the main aims of Chapter 1 is to draw a distinction between two related concerns. One goal of a study of word meaning must be to explicate the conditions under which a given word may apply to a given entity. Consequently, we may wish to ask when an entity may be said to be a *lemon*, that is, when the word *lemon* may apply to the entity in question. Often the answer is forthcoming: when that entity is a lemon. But one of the goals of the first chapter is to show that this does not follow. The issues of when an entity can be said to be an *X* and when that entity actually is an X are, *prima facie*, independent. The puzzles of Chapter 1 demonstrate this. The Lion Puzzle indicates the possibility that the same unambiguous word can be said both to apply and not to apply to one and the same entity: the word *lion* applies to a statue of a lion under one set of circumstances and does not apply under other circumstances. Such flexibility in the conditions under which *lion* is actually used cannot, however, be said to be due to an inherent flexibility in whether this entity actually is a lion. The question of whether a statue of a lion actually is a lion obtains an unequivocal negative answer. Similarly, in the case of Nunberg's Ham Sandwich. Here again we have a case where the use of a phrase, *ham sandwich*, crosses ontological boundaries. The cases of Lakoff's mother and Macken's father also provide a similar demonstration. Here too we can see that the issue of what constitutes an entity being a father or mother has only an indirect bearing on the use of the words *father* and *mother*.

The puzzles, then, are an attempt to show that we may choose not to conflate the ontological issue of what an X is, with the more obviously semantical issue of what conditions must hold for the word *X* to apply to an entity. Indeed, the rest of the thesis

has, in some way, been a statement of some of the benefits of rejecting this conflation. However, the puzzles demonstrate more than this. They point to the possibility that there appear to be different thoughts associated with the same linguistic string. That this is the case at all is unsurprising. Indexical expressions purport to convey different thoughts via the same linguistic expression. That this might be the case for expressions which are thought of as being neither indexical *nor* ambiguous, expressions such as *lion*, *father*, etc., is not only surprising but problematic. The problem centres on how it is that we may associate the same linguistic string with potentially indefinitely many different contents of thought.

A related problem in linguistics concerns apparent cases of non-intersective adjective noun combinations. A case such as *stone lion*, for example, appears not to be something that is both made of stone and a lion. Indeed, a *stone lion* is generally said not to be a real lion at all. But, this denies the inherent flexibility of word use. The word *lion* can indeed apply to stone statues of lions as long as the circumstances of use are of a non-standard kind. And it is by recognising this fact that we allow for the potential simplification of accounts of non-intersective adjective modification. It is a consequence of recognising the fact that *lion* may have the sense "statue of a lion" that the adjective-noun conjunction *stone lion* can be seen to be intersective. So our argument here was that, although there are cases of non-intersective adjectival modification, these in general may be related to the puzzles we have discussed and, further, that the proper treatment of the latter may lead to a considerably easier treatment of the former. Many of the complexities associated with analyses such as Levi's (1978) are not necessarily to do with the composition of adjective and noun but to do with the flexibility of one or other or both.

As we mentioned above, what is meant by flexibility is that different contents of thought may be associated with the same unambiguous linguistic expression. In the psychological literature, there are two areas upon which we may focus to gain insight into a psychological treatment of this flexibility of content.

One area concerns dictionary theories of mind. Herb Clark's work on sense creation sug-

gests that the same linguistic string may be associated with radically different thoughts. The problem, as he sees it, is that the same linguistic expression may be associated with infinitely many different thoughts. The problem, then, is that of deciding which of the many different senses of a word are stably mentally represented. The second area concerns the psychological literature on concepts. This literature, too, is concerned to detail not only the stable contents of semantic memory but what may be thought of as the stable "building blocks" of thought. So, here too, we may expect some kind of answer to the problem posed by the inherent flexibility of word uses. These, then, are the literatures we explored later in the thesis.

Another aspect of the first chapter was to introduce a further distinction, one that is as germane to psychological theories of sense as it is to psychological theories of concepts. Two claims are associated with these theories. One is a claim that the contents that we, as theorists, ascribe to concepts and senses, serves to classify behaviour in certain ways. The particular classificatory schema we adopt, however, must respect various constraints from the disciplines of cognitive science. That is, as far as a theory of word meaning is concerned, though all classifications are equal, in that they all classify, some classifications are more equal than others. This, then, allows us to choose between a number of different theories of concepts.

The other claim associated with theories of concepts and senses concerns mental representation. That is, concepts, with such and such content, are deemed to be mentally represented in a particular way. In this thesis we have made no detailed claims about mental representation. Indeed, it is hard to do so. For, as the theory of computation suggests, we can with no certainty progress from a claim about the *nature* of a function that is computed to a claim about the *way* that function is computed.

Having established a little theoretical framework, we then turned to some of the formal framework that we have employed throughout this thesis. The view we have adopted, taken directly from Situation Theory, is that word meanings are informational constraints relating types of uses of a word with certain worldly properties. We described the relation between conditional and unconditional constraints and attempted to fore-

stall the argument that the formulation of conditional constraints is incoherent by appealing to a general distinction between a semantical enterprise and a natural scientific one. In this chapter we tied the notions of sense and content, which we use for the rest of the thesis, to the situation theoretic notion of informational *content.* This was done to avoid confusion with existing notions rather than to tie us to a particular situation-theoretic approach. It is also here that we mentioned the connection between semantics and information, an association that allows, at least in principle, for a tying of semantic issues with epistemological ones. Developing the motivation for this general position would be both inappropriate and too difficult to do in this thesis. Instead, the position should be seen as just that: a position.

Finally we discussed the possibility of misinformation arising from concepts and we argued that this possibility does not constitute grounds for rejecting the position that concepts nonetheless express attunement to informational constraints. Thus armed, we then turned to some of the more obvious choices we might face in attempting to account for puzzles like those of Chapter 1.

In Chapter 2 we outlined some of the choices which must be faced by a cognitive scientific account of word meaning. The first choice concerns how we construe external meaning relations, relations between the types of use of a word and the properties about which those uses convey information. As we suggested in Chapter 1, the formal framework that we assume is that of Situation Theory, and this for reasons of its wedding of a theory of semantics to a theory of information. Another beneficial aspect of Situation Theory, however, is the choice it allows in analysing meaning relations in terms of informational constraints. Barwise & Perry (1983) talk, in particular, of the distinction between conditional and unconditional constraints and it is this choice that we considered first. We suggested that the prime consideration in making this choice was the treatment of defeasibility. Standardly, meaning relations may be seen in terms of entailment relations which exhibit the property of monotonicity or indefeasibility. Indeed, bar certain caveats, this is the picture we would assume were we to analyse meaning relations in terms of unconditional constraints. However, by treating meaning relations in terms of conditional constraints we allow the possibility that meaning relations may give rise

to defeasibility while not being themselves intrinsically defeasible. This, then, is the property of conditionality. Another property of constraints is there *a priori* nature. That is, given knowledge of such a constraint and knowledge of its "antecedent" situation type, knowledge of the "consequent" situation type is *a priori* in that it depends on no further empirical investigation. Consequently, a view of meaning relations which analyses them in terms of conditional constraints seems committed to a view of meaning relations as contingent *a priori* relations.

Another choice that may face an account of word meaning is whether we treat the various uses of words as resulting from attunement to a single constraint or to many constraints. Which choice we take will have a bearing on how we treat considerations of counterfactuals. If, for example, we treat the uses of a word in terms of a single constraint, then should we envisage counterfactual circumstances in which the word applies but in which the "antecedent" of the constraint is not satisfied, then we must re-examine our formulation of this single constraint. However, if we regard the uses of a word as belying multiple constraints then, were we to encounter the same problem, another solution would be available to us. Namely, we would claim that these counterfactual circumstances are ones in which *another* constraint held. And we could thus preserve our initial formulation of the first constraint.

Related to this is the third choice, that of whether we regard meaning relations as descriptional. That is, can the relata of such relations be given by description. Standardly, for certain words such as natural kind words, it is thought that meaning relations are not descriptional. But this position often rests on an argument from counterfactual circumstances. If we accept that meaning relations may be both conditional and that there are many of them underlying the uses of a single word, then, we argued, the claim that word meanings are descriptional may be maintained. That is, it is our position that the view that meaning relations are not descriptional is in some sense a corollary of the view that underlying the various uses of a word is a single indefeasible meaning relation.

Another set of choices we face in developing a cognitive scientific view of word meaning

concern what we assume language users to have as stable mental representations. We introduced the word LEXON to refer to the informational content that is stably mentally represented. The choices we faced are similar to those for the nature of meaning relations and concern, principally, the number of LEXONs we postulate for each unambiguous word, the content of these LEXONs and the nature of the relation between the contents of senses and LEXONs. The first choice concerns whether we wish to treat words which are strictly not ambiguous as either ambiguous or general, or whether we want to preserve the intuition that the content of the LEXON may express the default sense of the word. We suggested that, for the puzzles of Chapter 1, this last option was the more plausible. The second choice, the relation between the contents of senses and LEXONs, depends on the first. Assuming that senses can be quite specific in nature, the generality option would maintain that the contents of senses are but monotonic extensions of the content of LEXONs. An alternative, one that is required if LEXONs express the content of default senses, is to regard the content of senses as non-monotonically related to the content of LEXONs. The ambiguity option, depending on precise details, may well opt for one or other of these relations.

Finally, in Chapter 2, we considered a possible language, one based on Johnsons' (1988) attribute-value logic, in which we might express the contents of senses and LEXONs. This was for several reasons. Firstly, we found it necessary to be clear about the kinds of feature-structure that we employed later in the thesis. Secondly, we also found it necessary to be clear about the ontology assumed by such a language as this. Indeed, in Chapter 7, we had reason to question the efficacy of just such a language in describing the facts pertaining to concepts. And thirdly, we considered it important to distance ourselves, once more, from claims concerning the precise nature of mental representation. Our use of feature-structures, we pointed out, carried no necessary commitment to decompositional theories of meaning or to the claim that these features play a privileged role in mental computation. Rather, we see these feature-structures as simply classifying both linguistic behaviour and mental states. They do not represent claims about the precise nature of mental representation.

Having, in the first two chapters, pointed to some general issues for theories of word

meaning and sense, in Chapter 3, we turned to motivate a particular kind of theory of word meaning, Sense Generation. We began by considering the related theory of Sense Creation as developed by Herb Clark and his co-workers. Clark's claim is that non-conventional uses of words may have senses which are created for the nonce. His point is made by a consideration of so-called contextual expressions, expressions such as the denominal verb *to thesis*, which appear to have senses, or informational content, dependent solely on the circumstances of use. That is, the senses of contextual expressions have the properties of non-denumerability and contextuality. As Clark points out, such properties as these prove difficult to reconcile with traditional view of parsing, views which rest on the sense-selection assumption, the assumption that the senses of words may be finitely listed in a context independent manner. Now, though we agreed with much of Clark's argumentation, we disagreed on emphasis. In particular, we suggested that many expressions which Clark does not consider to be contextual expressions, in particular, indexical expressions and what he calls purely intentional expressions, may also exhibit the properties of contextuality and non-denumerability. And, further, and most importantly, we pointed out that Clark appears to restrict his attention to non-conventional uses of words. It has been our contention in this thesis that the conventional uses of words cannot be assimilated to the sense-selection view as Clark assumes.

With this caveat to Clark's views, we proceeded to consider two kinds of sense selection, strong and weak. Strong sense selection treats the various senses of unambiguous words as if they were the different senses of ambiguous words. That is, it fails to distinguish between ambiguity and degrees of vagueness. It also fails to respect the suggestion from Clark and others that the senses of a word may be infinite in number. Another form of sense selection, weak sense selection, we also found lacking as a psychological theory of sense. Its problems seemed twofold. Though it allowed for the possibility of treating unambiguous words as general, in the case of certain words, such as natural kind words, such an option seems to allow the possibility that the word may apply indiscriminately. That is, the satisfaction of the content of the corresponding LEXON would not, in general, prove sufficient for the application of the word. Avoiding the generality option, one is unfortunately forced to treat vagueness in terms of ambiguity

by postulating multiple LEXONs for the same unambiguous word. However, even were we to allow this to occur, there would be an equally serious problem with sense selection accounts. The argument comes from a consideration of the psychological literature on concepts and, in particular, the literature on conceptual coherence.

Sense selection accounts treat related senses as independent. Consider the two senses of *lion*: "real lion" and "statue of a lion". By postulating an independent LEXON for each sense, as sense selection accounts would, they fail to express the coherence or relatedness of their contents. We indicated that the arguments of Murphy & Medin, while perhaps in need of slight modification, are strongly suggestive that the use of words, being mediated by the application of concepts, is extremely sensitive to this coherence or relatedness. And, further, that a failure to capture this coherence must result in a failure to express either the psychology of concept use or the psychology of word use.

Having rejected some possible conceptions of sense, we turned to an exposition of Sense Generation. Sense Generation embodies the following assumptions as regards the choice of Chapter 2: it is assumed that there is one LEXON only for each unambiguous word; that the content of LEXONs expresses the default sense of the corresponding word; that the relation between senses and LEXONs is non-monotonic, the content of the sense being derived from that of the LEXON. In the remainder of the chapter, we described in more detail the properties of senses, words and LEXONs. Our main focus was, however, senses. In particular, we concentrated on relating the notion of sense which we had employed fairly intuitively in the early part of the thesis, to the notion introduced by Frege and expanded upon by Evans (1976). Part of our objective here was to articulate a criterion for individuating word senses and, perhaps more importantly, to re-orient our work towards the theme with which we began: the role of word meaning in an account of thought. In particular, by re-iterating Frege's original motivation for introducing the notion of sense, we made the link between senses and the thoughts of language users and, hence, of the concepts that mediate their uses of words.

In Chapter 4, we outlined a particular view of Sense Generation, rather unexcitingly

called the Relational View. Much of the motivation for the Relational View comes from the psychological literature on concepts. We argued for the importance of four particular morals in a theory of word meaning: the moral of coherence; the moral of representational economy; the moral of central exemplars; and the moral of context sensitivity. The rest of the chapter was taken up by an exposition of this Relational View in terms of Situation Theory and with an analysis of the puzzles. Our conclusions were: that this theory of word meaning respected the psychological literature on conceptual coherence; that it maintained simple economical representations for the senses of words; that it accorded central exemplars a privileged status as suggested by prototype theory; and that it respected some of Barsalou's (1987) considerations on the contextual sensitivity of the content of concepts.

At this point of the thesis our attention turned away from the psychological literature and towards a debate in the contemporary analytic philosophical literature concerning the "meaning" of proper names. The reasons for this apparent diversion were twofold. Firstly, to prepare the ground for a discussion of natural kinds in Chapter 6 which, in itself, was to prepare the ground for a discussion of prototype theory in Chapter 7. And, secondly, to relate the notion of meaning which we had so far employed, that in which meaning relations were construed as contingent *a priori* conditional constraints, to traditional notions of meaning. The literature on proper names is germane for another reason: it was this literature that first saw the introduction of the Fregean notion of sense only, then, to see its demise. Since we have tied the psychological conception of sense to the Fregean one, it was necessary to reflect on the reasons why Fregean senses have become unpopular.

So, in Chapter 5, we turned to the literature on proper names and, in particular, Kripke's argument against the possibility that proper names have Fregean senses. We began by outlining a revised version of Kripke's definition of description theories of proper names and by considering some of the motivations for description theories in general. In particular, they have been seen as providing accounts of existential statements, statements of identity and the determination of reference. Having prepared the ground, we then turned to the description theories of Frege and Russell, the first that Kripke criticises.

The early Frege held a theory of meaning in which semantic values are to be equated with Meaning or referents. He also held, however, that his theory of meaning should explain the cognitive significance of language. Thus, the theory should explain how it is that the same person may coherently take different attitudes to one and the same sentence at one and the same time. It should also explain the fact that statements of identity convey more information than the simple law of identity. How it is, for example, that *The evening star is the morning star* conveys information over and above the claim that an object is self-identical. However, if semantic values are equated with the objects of reference then, as Frege found, accounting for these facts is difficult. That is, there is a tension between the choice of semantic value and the goal of explaining cognitive significance. For reasons such as these, then, Frege rejected his earlier position in favour of a theory of meaning which distinguished between the Meaning, or referent, of a sentence and the thought or sense connected to it. And it is this notion of sense which, for Frege, was deemed to explain any possible cognitive difference between sentences having the same Meaning. Though Frege was inexplicit on this point, both Evans and Kripke seem committed to the view that Frege believed that senses can be given by description. As such, Frege's view falls directly under Kripke's rubric of description theories. And, though we will not re-iterate the points made earlier, the same is true of Russell's account of proper names.

Another category of theories of proper names is defined by what Kripke calls the "cluster-concept" approach. Roughly, the idea is that the meaning of a proper name can be given by a cluster of descriptions, some proportion of which must be true of the referent of that name. The accounts of Searle and Strawson, with the caveat that these are more properly thought of as theories of the determination of the referents of proper names rather than their meaning, both fall into this category. Such cluster-concept approaches are of interest for several reasons. First, they purport to explain how it is that language users associate different descriptions with the same name and yet achieve successful reference to the bearer of the name. Second, and perhaps more importantly given our interest in integrating facts about the epistemology of language users into semantic analysis, of the cluster of descriptions associated with a proper name, to each language user, a certain proportion are deemed to be known. That is, semantic facts

concerning proper names are explicated in terms of facts concerning the epistemological condition of language users. It is partly for this reason that our attention turned to Kripke's well-known and apparently successful attack on these description theories.

Kripke begins by outlining his conception of the analytic as the necessary *a priori* and, above all else, it is this conception which, for Kripke, prevents description theories from being theories of meaning. Necessity is primarily a metaphysical notion. It is a notion that concerns the way of this world and all other worlds: if a statement is true necessarily then it must be true in all possible worlds. That is, the world could not have been such as to make the statement false. Now, an explication of the analytic in terms of this notion, necessity, is what Kripke claims counts against the possibility of there being a description theory of the meaning of proper names. Supposing, for example, that *Aristotle* means *the author of De Anima*, then *Aristotle is the author of De Anima* would express an analytic truth. That is, it would express an *a priori* truth in that the truth of the statement would depend solely on the meanings of the terms of the sentence and not on any empirical investigation. And it would also express a necessary truth. But it is precisely this which Kripke disputes. *Aristotle is the author of De Anima* cannot express a necessary truth for it is not a necessary fact about Aristotle that he ever wrote De Anima. Aristotle might, presumably, never have put pen to paper, so to speak.

We indicated in Chapter 5 our general agreement with this argument but suggested several ways of concluding. One possibility is to attempt modifications to the descriptions associated with proper names in the hope that they may, then, express necessary truths. This possibility, adopted by cluster-concept approaches is one that is also rejected by Kripke. And, again, we are in agreement. A second possibility, the one adopted by Kripke, is to suggest that proper names simply do not have Fregean senses; their meanings cannot be explicated by description. And a third possibility is one that we adopted, namely, of treating the Kripkean argument as an inadvertent *reductio* on the conception of analyticity which it presumes. That is, our position is one diametrically opposed to Kripke's. While he jettisons Fregean senses in favour of the analytic as the necessary *a priori*, we jettison this view of the analytic in favour of Fregean senses.

In the last section of the chapter, we turned to developing an alternative account of analyticity. We began by noting that many of the intuitions behind the philosophical conception of analyticity do not explicitly tie the analytic to the metaphysical notion of necessity. A rejection of the analytic as the *necessary a priori*, then, should not be seen as constituting a rejection of the notion of analyticity, merely of its reformulation. With this in mind, we proceeded to demonstrate that the notion of meaning in situation theory, the notion we exploited in developing Sense Generation and the Relational View, suggests that the analytic might be construed as the contingent *a priori*. Though we have not developed the view in any detail, we did briefly illustrate how such a view allows one to preserve description theories of proper names.

In Chapter 6 our attention turned to natural kind terms and the issue of whether their meaning can be explicated in terms of descriptions. The issue is closely related to that of proper names and the development of the chapter proceeds in a similar manner. The difference is that, in this chapter, we considered not only philosophical theories but also psychological ones.

We began by considering the so-called classical theory of concepts, the theory which holds that a concept specifies necessary and sufficient conditions on entities to which the concept applies. The problem with applying this theory to the cases of natural kind terms is, quite simply, that there appear to be no necessary and sufficient conditions on the application of the corresponding concepts. Consequently, as a psychological theory of concepts, the classical approach has lost favour. And, in this, is a similarity with arguments in the philosophical literature. In particular, we drew the connection with the arguments of Kripke and Putnam against such views on word meaning as those of Katz & Postal (1964). These views similarly hold that the meaning of a natural kind term may be given by specifying necessary and sufficient conditions on its use. As both Kripke and Putnam point out, there appear to be no such conditions in the case of natural kind terms.

In the psychological literature, there has been a move away from classical theory and toward prototype theory, essentially a cluster-concept theory. Our position with regard

to prototype theory as an explication of the meaning of natural kind terms was that just as classical theory failed to find conditions which were truly necessary of entities to which the word applied, so prototype theory similarly fails. We argued that the only way in which this theory may allow for the expression of necessary truths was via the expression of tautologies which render the concept indiscriminate. That is, though it may express necessary conditions on word use, it fails to specify sufficient conditions. Again, we related this argument to Kripke's argument against cluster-concept approaches to proper names.

A third class of theory in the psychological literature goes by the name of hybrid or binary. These assume that a concept has two aspects, a conceptual core which is suited to description by classical theory and a set of identification procedures suited to description by prototype theory. Though such a distinction has much currency in the psychological literature, we suggested a number of problems. In particular, we suggested that the classical theory fails to account for conceptual cores for precisely the reasons it fails to account for complete concepts. Another problem with this proposal is, it seems, far more damaging. For, if the proposal is correct, then most of the empirical studies taken to motivate and support prototype theory must necessarily be construed as evidence concerning identification procedures and not conceptual cores. Since it is a claim of hybrid theories that it is the core and not the identification procedures which are implicated in thought, it follows that this empirical evidence has no bearing on the nature of thought. That is, most of the psychological studies of concepts can be taken to have had no relation to thought and it is this dreadful relegation of empirical studies of concepts which we must seek to avoid. Consequently, the hybrid proposal is one we found lacking.

However, the main problem for psychological theories of concepts is the same as for standard theories of word meaning, the problem of divining necessary and sufficient conditions. It is precisely this problem which allows Kripke and Putnam to reject the notion that the meanings of natural kind terms may be explicated in terms of descriptions. Our attention, then, shifted to the arguments of Kripke and Putnam. Having considered the proposal they make, we indicated a number of problems. One

was the assumption that there be essential properties possessed by all members of a kind. Another is the failure to furnish an explanation of how uses of natural kind terms can be informational. For it is generally the case that the presence of essential properties are not known to language users and so their presence or absence cannot determine the actual uses of natural kind terms. These uses must be made on some other basis, and it is this other basis which invests the uses with information, information concerning this basis. These problems, then, raise the possibility that the Kripke-Putnam position, based as it is on the conception of the analytic as the necessary *a priori*, may also be an inadvertent *reductio*. Indeed, given the position we adopted in the previous chapter, it is hardly surprising that this should be our conclusion. Consequently, we then turned to a brief re-iteration of the Relational View with respect to natural kind terms and suggested that in its workings it is not so far removed from Putnam's principal concerns. Namely, the treatment of exceptional cases.

We ended the chapter by pointing out some of the trends we have attempted to illustrate with respect to the psychological and philosophical literatures. In particular, the trend away from simple description theories and toward cluster-concept approaches. But also the trend of recent years of divergence of psychological from philosophical views, exemplified by the rejection by Kripke and Putnam of approaches so often advanced in the psychological literature. Our position is that much of this divergence and the trend towards cluster-concept approaches is forced by the conception of the analytic as the necessary *a priori*. Our hope is that by adopting the position in which the analytic is equated with the contingent *a priori*, a reconciliation between the concerns of these psychologists and philosophers will be forthcoming.

In our last main chapter, Chapter 7, our attention turned away from the philosophical literature and towards the psychological literature on concepts, in particular, towards prototype theory.

Prototype theory stands as a particularly robust and popular statement of what a concept is. We began, then, by specifying the fundamental aspects of prototype theory: its commitment to describing concepts in terms of clusters of weighted attribute-

value descriptions; similarity being determined by attribute-value matching and the size of weights; typicality of exemplars being determined by similarity; and there being a threshold on similarity. To further exemplify the sort of theory which forms the focus of our attention, we offered two explicit accounts of prototype concepts, those of Cohen & Murphy (1984) and Smith et. al. (1988). These we then classified in terms of the choices outlined in Chapter 2. That is, we argued that prototype concepts thus construed can only sensibly be seen as expressing attunement to unconditional constraints and, further, that each prototype concept expresses attunement to only one such constraint.

Having attempted a classification of prototype theory, our attention turned to some of its more obvious problems: the problems of accounting for conceptual coherence; for developmental adequacy; representational economy; central exemplars; and contextual sensitivity. It is this family of problems, indeed, that motivates the account that we develop in the following section, the Family of Constraints view, formerly, the rather unexcitingly labelled Relational View. After a brief exposition of the view, we attempt to furnish explanations of how the problems faced by prototype theory are resolved by the Family of Constraints View. This done, we attempt the major goal of this chapter, which is to attempt a detailed comparison of a prototype view of concepts with the one we have developed.

We begin by outlining the Family of Constraints View in its barest essentials. The first problem of comparison is that prototype concepts express attunement to single unconditional constraints, whereas, under our account, underlying the use of a word are multiple conditional constraints. Our first move, then, is to reduce our multiple conditional constraints to one unconditional constraint. First we are required to conditionalise the conditional constraints. Though there are good reasons to believe such a move is not valid it is one we must take in order to compare our view with prototype theory. Conditionalisation, then, makes our multiple conditional constraints unconditional. The next step is to abstract over the combining concepts which are implicated in these unconditional constraints and which we argued play such an important role in determining conceptual coherence. Our third step is to abstract over the remaining background conditions which now figure in the antecedent condition of our now

unconditional constraints. It was the presence of these background conditions which we suggested gave an account of the context-sensitivity of concept and word use. The fourth step requires us to combine the constraints with which we are left, constraints which are unconditional, and make no reference to other concepts or background conditions. The pitfall associated with this particular move is that, in reducing the number of constraints which these different theories of concepts assume underly the use of these concepts, we make an ontological commitment. That is, the difference between assuming that there is one constraint instead of, say, ten is the same as that between saying that in a room there is one chair instead of ten. Though in the case of constraints, we may have no privileged knowledge of the actual number of constraints, it is clear that each theory makes a different claim about the nature of reality, just as if each disputed the actual numbers of chairs in a room. The fifth step requires us to ignore certain relations between attributes, namely, the possibility that in the FoC view we can encode the fact that certain attributes must co-occur, that certain other attributes cannot co-occur. This information is lost in formulating prototype theory. The final step is to interpret the redundancy in the constraint that we have now formulated in terms of a weighting on attribute-values. But, as we indicated, such a step can only make sense if one is committed to the notion that concepts express abstractions across all contexts. These moves, then, serve to convince us that the FoC View we have developed is better equipped to account for what is known of the psychology of concepts than prototype theory.

Our parting shot in this chapter was to indicate that the Family of Constraints View is also better placed to account for Wittgenstein's much quoted observations on family resemblance. In particular, the account we have developed illustrates how one can account for his notion of family resemblance while respecting his intuition that the fact that there may be quite disparate uses (either of words or concepts) does not indicate that the word or concept is disjunctive. In a similar vein, we also indicated that the account we have offered, while preserving the notion that there are stable mental representations underlying concepts, is entirely consistent with Barsalou's claims that concepts are nought but "analytic fictions".

8.2 Implications and Future Research

The main implications of the research presented in this thesis divide naturally into five distinct categories. The first concerns some empirical predictions of the FoC view we developed in Chapter 4.

8.2.1 Empirical Psychology

The view of word meaning that we have developed in this thesis suggests a number of empirical studies which fall, roughly, into two categories. The first concerns the psychology of language acquisition, the second the relation between psychological views of proper names and natural kind terms.

The FoC view, then, has implications for the precise course of acquisition of words relevant to some domain. There are two kinds of acquisition that may be studied. One concerns the acquisition of words by children, the other concerns the acquisition of words by novices in some, more technical, learning experience.

Children routinely experience difficulties in learning to which entities a word may apply. They may over-extend as, for example, when a child uses "moon" to refer to all round objects. Or they may under-extend as, for example, when a child uses "shoes" only to refer to a particular pair of shoes in a particular location. Now, prototype theory seems to predict a similar sort of underextension. Namely, prototype theory seems committed to the view that there is no generalisation of exceptional cases across categories. For example, a child's discovering that lemons can be squashed seems a sufficient learning experience for the child to realise that many things can be squashed, not simply lemons. However, the suggestion from the various models of prototype concepts is that this is not the case, that, in fact, such an experience will only change our concept of *lemon*.

The FoC view, however, predicts that such a learning experience can concomitantly alter a wide number of different concepts and so affect the use of a comparable number

of words. The suggestion is that learning that a squashed lemon can be called a lemon may be sufficient for a child to learn that the word *orange* may similarly be applied to squashed instances of oranges even if these have not formed a part of the child's experiences. So one implication of this research is that there may be empirical grounds for distinguishing the FoC view from the models of prototype concepts we have considered. One possibility is for a small series of experiments with children of various ages. The aim would be to investigate changes in the uses of already learned words, belonging to a certain semantic field, to apply to particular presented objects, and what effect the learning of new words in the semantic field has on the uses of the previously learned words. The prediction being that learning of new words, particularly those that label operations, can, in certain contexts, widen the domain of application of previously learned words. For this reason, it may be more appropriate to arrange the experiments in the form of a computer game where objects which play a role in the game can be freely manipulated. The goal would be to investigate whether children would both comprehend, and produce, uses of words to describe entities that would not normally be described by those words. And, further, whether these controlled changes in word use also affected the uses of other words.

A second implication for language acquisition concerns the acquisition of "expert" language: for example, the change in language use that accompanies the transition from novice to expert in some particular domain. An example of such a domain is medical diagnosis where it is of considerable importance that the right word is used to label the right condition. Similar to the study outlined above, one possibility would be to model how, for example, the use of terms for certain medical conditions shifted as diagnosticians became more expert. So, for example, how the use of the word *epilepsy* might change as subjects progressed from medical students to expert consultants. Another aspect of this work might be to develop a computational treatment, possibly with a view to interfacing this with existing expert systems in the area of medical diagnosis.

Another possible implication of the FoC view concerns the relation between psychological approaches to concepts and psychological approaches to proper names and face recognition. Whereas the arguments of Kripke and Putnam suggest that proper names

and natural kind terms are quite alike, recent work in the psychological literature conducted by Gillian Cohen suggest that these terms can be distinguished by their psychological properties. So, one aspect that could be considered is whether a Sense Generation approach to proper names would receive psychological support.

8.2.2 Linguistic Coverage

One interesting issue relating to the research we have described concerns the possibility of assimilating work done on tense and aspect with the FoC view. The FoC view is committed to the fact that the different senses of words may be non-monotonically related to the lexical entries of LEXONs from which they are generated. Recent work on tense and aspect (Moens & Steedman, 1988) also seems committed to such a non-monotonicity, *type coercion.* In *Max ran*, for example, "run" may be assigned, by default, the *process* aspectual type. However, in certain circumstances, *Max ran* may exhibit a *culminated process* reading for "run". As, for example, where one knows that Max runs one mile every day. The way this has been analysed is by claiming that the aspectual type has been coerced into another, the relation between the types being non-monotonic. It seems that, *prima facie*, these two different approaches relying, as they do, on similar solutions can be assimilated. This is of great interest since Sense Generation approaches have largely been applied only to noun phrases and this due to their relatively simple inflectional morphology. However, the possibility of an integration raises even more interesting questions. In the example with *run*, above, it seems that, aside from the change in aspectual type, no other semantic change occurs. But, presumably, these type coercion principles cannot apply universally. For example, there may be some verbs for which there cannot be a *culminated process* reading. The interest for the FoC view, here, is whether in such cases the coercion principles can be said to apply but only if there is an accompanying change in sense of the main verb. So we have not only aspectual change but other semantic change too. Isolating such cases would indeed be suggestive that the FoC view is touching on a central aspect of language.

8.2.3 Computational Concerns

Though we have not illustrated the relevant programs here, part of the strategy behind the research reported here was the writing of two short prolog programs to illustrate the FoC view of word meaning. One program attempts to demonstrate how the same word may be associated with different conditions of use, these new conditions being related to the old by some process of concept combination. Another program attempts to demonstrate how the interpretive indeterminacy of complex nominals such as *coal merchant* may be explained in terms of an indeterminacy in either unifying the descriptions associated with the component words, or in searching for an appropriate concept which relates these descriptions. Both programs employ standard feature structure descriptions to describe the contents of lexical entries. Both programs also appear to raise questions concerning the possible integration of psychological facts concerning word meaning and unification-based approaches to grammar. So one possibility which may be of interest is that the FoC view we have developed may be cast in terms of some or other computational linguistic theory: CUG, UCG, HPSG, etc.

Another aspect which is of interest concerns the appropriateness of a language such as an attribute-value language to describe facts concerning the psychology of word use. In particular, the work of Erickson & Mattson (1981) suggests that, even were we to be successful in associating feature structures with word senses and in combining them according to the syntax of some associated sentence, there would still be a deficit. Erickson & Mattson showed that people may parse sentences apparently successfully even when these sentences contain clear semantic contradictions. For example, the majority of their subjects, when asked *How many animals did Moses take into the Ark?*, would answer two. Even though *Moses* took none. Seemingly, listeners and parsers need to be capable, at least, of remaining insensitive to such contradictions. Unification, strictly, cannot be so insensitive. The interest is firstly what kind of operation we require over the content of lexical entries and secondly how it relates to standard unification. What this consideration raises is the possibility of exploring modifications to the standard unification operation over feature-structures and, indeed, of possible modifications to attribute-value languages. Thus, one might want to consider the possibility of what

we might call *partial unification.* This would involve unifying feature structures at their top-most level, that is, at the level of atomic values, ignoring embedded feature structures. This, at first glance, offers the possibility of a unification-like operation which preserves insensitivity to contradictions. Such thoughts, though, are clearly at an embryonic stage.

8.2.4 Formal Aspects

Sense Generation approaches make the key assumption that the generation of senses is essentially a non-monotonic process. That is, a sense generated from a LEXON may not be subsumed by that LEXON. This relates Sense Generation approaches and the FoC view quite naturally to a large literature in Artificial Intelligence on Knowledge Representation (KR). Most approaches to KR take as fundamental the distinction between default and exceptional cases of categories. For example, default birds fly but exceptional ones, such as penguins, do not. The important issue for cognitive science is that concerning exactly which knowledge we represent, and how we represent it, in capturing the facts of everyday reasoning. The standard approach has been to assume that people reason with defaults and only in cases where this leads to inconsistency do we override default properties and infer exceptional ones. The resulting "logic" of everyday reasoning is then non-monotonic in that conclusions may have to be retracted as further premisses are added. The rejection of this central tenet, namely monotonicity, of classical logic has led Israel (1980) among others to claim that non-monotonic logic is, quite simply, a contradiction in terms. Though there are other objections to such logics, one way of following up on the research reported in tis thesis is to concentrate on psychological evidence which suggests that such logics are unlikely to capture the facts of human reasoning. In particular, the facts which underly the psychology of word meaning.

Now, Sense Generation approaches assume that one particular sense will be used by default but that where this default sense leads to inconsistencies a different sense will be generated. The FoC view we have developed is of great interest here, principally

because we have couched it in terms of Situation Theory. In fact, when we phrased the relation between default and generated senses in terms of Situation Theory we did not get the same picture of non-monotonicity that we do in the standard AI formalisations. In particular, we get a system whose properties are best described not in terms of a single logic which is non-monotonic but in terms of several different logics. Each logic being defined by the set of background conditions to which it is relativised. The interest, then, for this approach to word meaning is the precise correspondence between a situation-theoretic account of Sense Generation and one which is couched in terms of non-monotonic logic. An interesting investigation, then, might pursue the divergences, and parallels, between a situation-theoretic account and more standard AI accounts of what is seen as the fundamental defeasibility associated with word meaning.

Another aspect of this particular implication of the FoC view is concerned with the relation between this situation-theoretic model of apparent non-monotonicity and psychological evidence. An example of a constraint which empirical psychological studies may place on the account of non-monotonicity we may hope to develop comes from an experiment performed by Erickson & Mattson (1981). In this they found that the standard assumption that the complete sense of a word is accessed during parsing does not always hold. Indeed, only partial information concerning a word's sense may be retrieved, the degree of retrieval being seemingly dependent on linguistic and non-linguistic context. This has important ramifications for work in Knowledge Representation where it is assumed that knowledge, say of the meaning of a word, comes in complete packets which are either accessed wholesale or not at all. Thus, what this experiment suggests is that a more subtle view of knowledge representation is required in which the degree of accessing of knowledge is dependent on contextual factors among others. Such findings as these, it seems, must be integrated with work on non-monotonic logics and situation-theoretic alternatives if we are to achieve a formal understanding of the psychology of defeasible reasoning.

8.2.5 Philosophical Concerns

The philosophical aspects of this thesis research concern, principally, an evaluation of Situation Theory in the light of arguments from the traditional analytic philosophical literature. They divide into three categories. The first concerns accounts of vagueness; the second concerns the significance of situation theory's notion of conditional constraint; and the third concerns different types of realism.

Vagueness

Traditional formal approaches to word meaning have assumed that words are either ambiguous or that they have a fixed sense which can be selected from the lexicon. Such assumptions are not, in general, valid as Clark (1983) has argued. However, it is assumed that, however vague relations are to be treated, these assumptions nonetheless hold of them. The FoC view has been motivated by the opposite conclusion: vagueness is not due to fuzziness or somesuch of senses but to the indeterminacy inherent in deciding which sense a word has on any particular occasion of use. It is not that the senses of words apply fuzzily to various objects, on the contrary they either apply or not. Vagueness, rather, should be seen as arising from the fact that a word may have many different senses and what is negotiable is not whether a given sense applies to some object, but which sense the word has. Such a view allows solutions to a number of problems associated with vagueness as well as metonymy and some degree of metaphor. However, what such a view requires is a radically different picture of meaning.

Traditional theories of meaning hold that the meaning of terms such as *lemon* must be explicated in terms of necessary and sufficient conditions for being a lemon. Accordingly, traditional theories of concepts and of the senses of words hold that, these express information about the nature of the objects to which they apply. The problem for most of these approaches is that such information may not be within the grasp of language users. Certainly, most language users have no, even implicit, knowledge of the necessary and sufficient conditions for lemonhood, even though, in most respects, they may appear

to have grasped the meaning of *lemon*. Sense Generation is an attempt to explicate the senses of words and the concepts which may underly their use in terms of properties which are not necessarily singly necessary and jointly sufficient. The issue is, to what extent are such senses or concepts anything to do with *meaning*? After all, the traditional theories argue, meaning is to do with necessity; the kinds of senses which psychological arguments discuss are to do with properties that are contingent. This, then, is where we can see the importance of conditional constraints. They reconcile the notions of contingency and meaning. So the senses which, according to Sense Generation, underly the uses of words can indeed be seen as expressing some aspect of the meanings of those words. This allows for a reconciliation between psychological and philosophical conceptions of meaning which was strictly disallowed under traditional theories of meaning. This, it seems, is an important aspect of Situation Theory. Another important aspect of Situation Theory is that it can be applied via Sense Generation to vague predicates.

Vagueness has dogged formal accounts of word meaning. The vagueness of relations such as "father", for example, has been a challenge to frameworks which are capable, at the lexical level, of only expressing ambiguity. Different formal proposals have been made and have generally been rejected. Osherson & Smith (1981) argued strongly that fuzzy set theory could not accommodate the facts pertaining to the compositionality of certain adjective-noun conjunctions. Similarly, Kamp & Partee (1989) cast doubt on the possibility of the technique of supervaluations being able to account for compounds such as *stone lion*. In this thesis, however, we have argued that the Relational View in particular and Sense Generation in general *can* account for expressions such as these. Indeed, this thesis is an attempt to account for these in a way which not only respects the cognitive sciences but which is also amenable to a formal account. The apparatus of Situation Theory makes such an account possible. The fact that it makes possible a formal account of vagueness seems to be an important result of Situation Theory.

Another account of the vagueness of relations such as "father" is that of Betsy Macken (forthcoming). An interesting question that is raised by the contrast between our accounts, is the issue of the *projection* of the argument roles of relations. Currently, this is the mechanism by which Barwise has analysed perspectival relations. Interesting ques-

tions to pursue here are: are these accounts mere "notational variants" of one another: are there two kinds of perspectival-relativity associated with relations which require these two accounts: can the projection analysis account for the range of cases which motivate Sense Generation? Seeking answers to questions such as these would appear to be an interesting and significant enterprise.

Conditional Constraints and Analyticity

Kripke's seminal article *Naming and Necessity* amounts to a vigorous critique of description theories of meaning, both of proper names and natural kinds. His conclusions are familiar: that proper names and natural kinds cannot be said to have senses, merely meanings. That is, merely referents, designata, denotations. For Kripke, such terms *rigidly designate.* His argument, and the similar arguments of Putnam (1975), however, rely heavily on the traditional conception of analyticity. This conceives of the analytic as being the necessary *a priori.* It is the focus on necessity which is of interest here. For Situation theory's notion of conditional constraint is such as to suggest that meaning may well be explicated in terms of statements which are merely *contingently* true. If indeed we view some, though perhaps not all, analytic truths as being only contingently true, then the Kripkean position is severely weakened. Indeed, one interpretation of his argument is as a *reductio* on the traditional conception of analyticity. However, it is important to realise that merely allowing contingency to be a facet of meaning does not totally undermine Kripke's position. For if we assume that proper names and natural kind terms have only one meaning then we cannot escape the consequence that our contingent explications of their meaning will not apply in all situations in which the words themselves can nonetheless be used. Such a fact allows Pratt (1987) to argue that constraints, even conditional ones, cannot explicate meaning. However, if we reject this assumption, the Unitary assumption, that such terms have only one meaning underlying their uses then Pratt's objections as well as the Kripke-Putnam position is undermined.

There are a number of consequences that follow from this undermining. One is quite

simply that description theories of both proper names and natural kind terms cannot be discounted as theories of meaning by *fiat.* Another is that we can begin a reconciliation of philosophical and psychological approaches to word meaning. Still further, we have effected a certain "cleaving" of semantics from ontology. This can be seen in a number of ways. The clearest manifestation of it is in the fact that we have weakened the ontological commitment made by traditional theories of meaning. In particular, these rested on a commitment to the existence of *necessary* truths. We now simply require there to be contingent truths, a much weaker ontological commitment. However, the real force of this "cleaving" is when we consider the relation between realist and anti-realist semantic accounts.

Ontological Realism and Semantic Anti-Realism

Situation Theory is a strongly realist theory, but the force of its realism is ontological. In this there is no conflict with the account of word meaning I have been developing. An aspect of semantic realism is not only this ontological commitment, but a semantic one. Statements are deemed to have a determinate truth value, even statements which must necessarily transcend our knowledge. Statements, for example, about the past, the future and about other minds. The point about such statements is that their truth necessarily transcends our knowledge and so grasping their meaning must consist of something other than apprehending their truth or falsity. But grasp their meaning we surely do. Rejection of this tenet of semantic realism is, however, something which is perfectly compatible with Situation Theory.

Wittgenstein's notion of criteria can be seen as an alternative to this strong claim associated with semantic realism. Grasping the meaning of a term is to be equated with knowing the conditions under which the use of that term is *warranted.* These criterial relations, though conferring certainty, are weaker than entailment relations, indeed they are defeasible. So criterial relations avoid the claims of semantic realism: to make an ascription of pain to another individual is not to know the circumstances which make such an ascription true, for these are beyond our grasp. Rather, making such

an ascription consists in our knowing what circumstances warrant such an ascription, circumstances which are within our grasp, such as howling, going red in the face, jumping up and down, etc. Such relations can be defeated, of course, but such is their nature. To see that Situation Theory offers ontological realism without a commitment to this strong tenet of semantic realism, we need only recognise that conditional constraints have the same characteristics as criteria relations. This is the subject of Braisby & Franks (in preparation). This fact of Situation Theory adds to its philosophical importance. It also makes it a more obviously attractive proposition for the cognitive sciences.

Bibliography

Armstrong S. L., Gleitman L. & Gleitman H. (1983). What some Concepts might not be. *Cognition, 13,* 263–308.

Baker (1974). Criteria: a New Foundation for Semantics. *Ratio, 16,* 156–189.

Barsalou L. W. (1982). Context-Independent and Context-Dependent Information in Concepts. *Memory & Cognition, 10,* 82–93.

Barsalou L. W. (1985). Ideals, Central Tendency and Frequency of Instantiation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 211–227.

Barsalou L. W. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In Neisser U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 101–140. Cambridge: Cambridge University Press.

Barsalou L. W. & Sewell D. R. (1984). *Constructing Representations of Categories from Different Points of View.* Emory Cognition Project Report No. 2. Emory University, Atlanta, GA.

Barwise J. (1985). *The Situation in Logic-II: Conditionals and Conditional Information.* CSLI Report No. CSLI–85–21.

Barwise, J. (1989). *The Situation in Logic.* CSLI Lecture Notes 17. University of Chicago Press: Chicago, Illinois.

Barwise J. & J. Perry (1983). *Situations and Attitudes.* Bradford Books, MIT Press.

Braisby N. R. (1989a). Situating Word Meaning. In Braisby, N. R. & Cooper R. P. (eds.) (1989). *Situation Theoretic Studies in Psychology, Language and Logic.* Edinburgh Working Papers in Cognitive Science, Volume 3. Centre for Cognitive Science, University of Edinburgh. September 1989.

Forthcoming in Cooper R., Mukai K. & Perry J. (eds.) (1990). *Situation Theory and its Applications, I.* Proceedings of the First Conference on Situation Theory and its Applications.

Braisby N. R. (1989b). Prototypes and Word Meaning. Talk given at the First Edinburgh Round Table on the Mental Lexicon. Edinburgh, UK. June 1989.

In Dunbar G. L., Franks B. W. & Myers T. F. (eds.) (1989). *Papers from the Edinburgh Round Table on the Mental Lexicon.* Edinburgh Working Papers in Cognitive Science, Volume 4. Centre for Cognitive Science, University of Edinburgh. January 1990.

Braisby N. R. (forthcoming). Some Puzzles of Word Meaning and a Computational Treatment.

Braisby N. R. & Franks B. W. In preparation. Criteria and Conditional Constraints.

Brown E. K. & Miller J. E. (1980). Syntax: A Linguistic Introduction to Sentence Structure. London: Hutchinson.

Carey S. (1982). Semantic Development: the State of the Art. In Wanner E. & Gleitman L. R. (eds.) *Language Acquisition: The State of the Art,* 347–389. Cambridge: Cambridge University Press.

Clark, H. H. (1983). Making Sense of Nonce Sense. In d'Arcais, G. B. F. and Jarvella, R. J. (eds.). *The Process of Language Understanding.* Chichester: John Wiley and Sons. 297–331.

Clark H. H. (1989). Words, the World, and their Possibilities. Manuscript. Stanford University.

Clark H. H. & Gerrig R. J. (1983). Understanding Old Words with New Meanings. *Journal of Verbal Learning and verbal Behaviour, 22,* 591–608.

Cohen B. & Murphy G. L. (1984). Models of Concepts. *Cognitive Science, 8,* 27–58.

Cohen G. (1983). *The Psychology of Cognition.* London: Academic Press.

Cresswell M. J. (1978). Semantic Competence. In Guenthner F. & Guenthner-Reutter M. (eds.) *Meaning and Translation: Philosophical and Linguistic Approaches,* 9–27. London: Duckworth.

Cruse D. A. (1986). *Lexical Semantics.* Cambridge: Cambridge University Press.

Erickson T. B. & Mattson M. E. (1981). From Words to Meanings: a Semantic Illusion. *Journal of Verbal Learning and Verbal Behaviour, 20,* 540–551.

Evans G. (1976). *The Varieties of Reference.* Oxford: Basil Blackwell.

Fillmore C. (1982). Towards a Descriptive Framework for Spatial Deixis. In Jarvella R. J. & W. Klein (eds.). *Speech, Place and Action: Studies in Deixis and Related Topics.* Chichester: John Wiley and Sons. 31–59.

Fodor J. A. Garrett M. F. Walker E. C. T. & C. H. Parkes (1980). Against Definitions. *Cognition, 8,* 263–367.

Franks B. W. (1989). Criteria and Concepts: an Anti-Realist Approach to Word Meaning. PhD. Thesis. Centre for Cognitive Science, University of Edinburgh.

Franks B. W. & Braisby N. R. In Preparation. What is Sense Generation?

Franks B. W. & Braisby N. R. (1990). Sense Generation or How to Make a Mental Lexicon Flexible. Paper submitted to the 12^{th} Annual Conference of the Cognitive Science Society, MIT, July 1990.

Franks B., Myers T. & McGlashan S. (1988). Defeasibility in Concept Combination. In *Proceedings of the Tenth Annual Conference of the American Cognitive Science Society,* Montreal.

Frye D., Braisby N. R., Lowe J., Maroudas C. & Nicholls J. (1989). Young Children's Understanding of Counting and Cardinality. *Child Development, 60,* 1158–1171.

Grice H. P. & P. F. Strawson (1956). In Defense of a Dogma. *The Philosophical Review, 65.*

Hampton J. A. (1990). *Psychological Models of Concepts.* Talk given at the Conference on Concepts and Categorisation, University of Sussex. April, 1990.

Hull C. L. (1920). Quantitative Aspects of the Evolution of Concepts. *Psychological Monographs, 28.*

Israel D. (1980). What's wrong with Non-monotonic Logic? In Proceedings of the American Association for Artificial Intelligence, 1980.

Johnson M. (1988). *Attribute-Value Logic and the Theory of Grammar.* CSLI Lecture Notes 16. University of Chicago Press: Chicago, Illinois.

Johnson-Laird P. N. (1987). The Mental Representation of the Meaning of Words. *Cognition, 25,* 189–211.

Johnson-Laird P. N. & Wason P. C. (eds.) (1977). *Thinking: Readings in Cognitive Science.* Cambridge: Cambridge University Press.

Kamp H. & Partee B. (1989). Prototype Theory and Compositionality. Unpublished ms. University of Stuttgart.

Katz J. J. & Postal P. (1964). *An Integrated Theory of Linguistic Descriptions.* Cambridge, Mass., MIT Press.

Kay P. & Zimmer K. (1976). On the Semantics of Compounds and Genitives in English. Paper presented at the Sixth Annual Meeting of the California Linguistics Association. San Diego, California.

Kripke S. A. (1972). Naming and Necessity. In Davidson D. & G. Harman (eds.). *Semantics of Natural Languages.* Dordrecht: Reidel.

Lakoff G. (1987). Cognitive Models and Prototype Theory. In Neisser U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 63–100. Cambridge: Cambridge University Press.

Lakoff G. (1987). *Women, Fire and Dangerous Things: What Categories reveal about the Mind.* Chicago, Illinois: The University of Chicago Press.

Levi, J. N. (1978). *The Syntax and Semantics of Complex Nominals.* Academic Press.

Lewis D. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy, 65,* 113–126.

Lyons J. (1968). An Introduction to Theoretical Linguistics. Cambridge University Press.

Macken B. (forthcoming). Out of the Mouths of Babes. Forthcoming in Cooper R., Mukai K. & Perry J. (eds.) (1990). *Situation Theory and its Applications, I.* Proceedings of the First Conference on Situation Theory and its Applications.

McDowell J. (1982). Criteria, Defeasibility and Meaning. In Proceeding of the British Academy.

Medin D. L. & Wattenmaker W. D. (1987). Category Cohesiveness, Theories and Cognitive Archeology. In Neisser U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 25–62. Cambridge: Cambridge University Press.

Miller G. A. & Johnson–Laird P. N. (1976). *Language and Perception.* Cambridge: Harvard University Press.

Moens M. & Steedman M. J. (1988). Temporal Ontology and Temporal Reference. *Computational Linguistics, 14,* 15–28.

Morrow D. G. & Clark H. H. (1988). Interpreting Words in Spatial Descriptions. *Language and Cognitive Processes, 3(4),* 275–291.

Murphy G. L. & Medin D. L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review, 92,* 289–316.

Myers, T. F. B. W. Franks & N. Braisby (1989). Partiality and Coherence in Concept Combination. Forthcoming in the Proceedings of the International Colloquium on Cognitive Science, San Sebastian, Spain. May 1989.

Nunberg G. D. (1978). *The Pragmatics of Reference.* PhD Thesis. Reproduced by the Indiana University Linguistics Club.

Osherson D. N. & Smith E. E. (1981). On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition, 9,* 35–58.

Phillips C. (1978). Constructivism and Epistemology. *Journal of Philosophy, 53,* 51–69.

Platts, M. (1979). *Ways of Meaning.* London: Routledge and Kegan Paul.

Pratt I. (1987). Constraints, Meaning and Information. *Linguistics and Philosophy, 10,* 299–324.

Putnam H. (1975). Is Semantics Possible? Chapter 8 in *Mind, Language and Reality, Volume 2: Philosophical Papers.* Cambridge: Cambridge University Press.

Putnam H. (1975). The Meaning of 'Meaning'. Chapter 8 in *Mind, Language and Reality, Volume 2: Philosophical Papers.* Cambridge: Cambridge University Press.

Putnam H. (1978). Meaning, Reference and Stereotypes. In Guenthner F. & Guenthner-Reutter M. (eds.) *Meaning and Translation: Philosophical and Linguistic Approaches,* 61–81. London: Duckworth.

Quine W. V. O. (1948). On What There Is. *Review of Metaphysics.*

Quine W. V. O. (1951). Two Dogmas of Empiricism. *The Philosophical Review, 60.*

Quine W. V. O. (1960). *Word and Object.* Cambridge, Mass.: MIT Press.

Rosch E. H. (1973). On the Internal Structure of Perceptual and Semantic Categories. In Moore T. E. (ed.) *Cognitive Development and the Acquisition of Language.* New York: Academic Press.

Rosch E. H. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General, 104,* 192–233.

Rosch E. H. & Mervis C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology, 7,* 573–605.

Rosch E. H., Mervis C. B., Gray, W., Johnson D. & Boyes-Braem P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology, 8,* 382–439.

Roth E. M. & Shoben E. J. (1983). The Effect of Context on the Structure of Categories. *Cognitive Psychology, 15,* 346–378.

Russell B. (1956). On Denoting. In March R. C. (ed.) *Logic and Knowledge.* London: Allen and Unwin. 41–56.

Searle J. R. (1958). Proper Names. *Mind, 67,* 166–73.

Seligman J. (1989). Perspectives in Situation Theory. Paper presented at the Conference on Situation Theory and its Applications, Asilomar, California. March, 1989.

Shieber S. M. (1986). *An Introduction to Unification-Based Approaches to Grammar.* CSLI Lecture Notes 4. University of Chicago Press: Chicago, Illinois.

Shieber S. M., Uszkoreit H., Pereira F. C. N., Robinson J. J. & Tyson M. (1983). The Formalism and Implementation of PATR-II. In *Research on Interactive Acquisition and Use of Knowledge, Artificial Intelligence Center,* SRI International, Menlo Park, California.

Smith E. E. & Medin D. L. (1981). *Categories and Concepts.* Cambridge: Harvard University Press.

Smith E. E. & Osherson D. N. (1984). Conceptual Combination with Prototype Concepts. *Cognitive Science, 8,* 337–361.

Smith E. E., Osherson D. N., Rips L. J. & Keane M. (1988). Combining Prototypes: A Selective Modification Model. *Cognitive Science, 12,* 485–527.

Strawson P. F. (1959). *Individuals.* Methuen, London.

Tversky A. (1977). Features of Similarity. *Psychological Review, 84,* 327–352.

Vygotsky L. S. (1962). *Thought and Language.* MIT Press.

White M. (1950). The Analytic and the Synthetic: An Untenable Dualism. In *John Dewey: Philosopher of Science and Freedom.* New York, The Dial press.

Whitehead Z. N. & Russell B. (1927). *Principia Mathematica,* Vol. 1. CUP: Cambridge.

Wittgenstein L. (1953). *Philosophical Investigations.* Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.

Zadeh L. (1965). Fuzzy Sets. *Inform. Control, 8,* 338–353.

Zwicky A. M. & Sadock J. M. (1975). Ambiguity Tests and How to Fail Them. In Kimball J. P. (ed.) *Syntax and Semantics, 4*, 1–36. Academic Press.

Appendix: Publications

Braisby N. R. (1989a). Situating Word Meaning. In Braisby, N. R. & Cooper R. P. (eds.) (1989). *Situation Theoretic Studies in Psychology, Language and Logic.* Edinburgh Working Papers in Cognitive Science, Volume 3. Centre for Cognitive Science, University of Edinburgh. September 1989.

Forthcoming in Cooper R., Mukai K. & Perry J. (eds.) (1990). *Situation Theory and its Applications, I.* Proceedings of the First Conference on Situation Theory and its Applications.

Braisby N. R. (1989b). Prototypes and Word Meaning. Talk given at the First Edinburgh Round Table on the Mental Lexicon. Edinburgh, UK. June 1989.

In Dunbar G. L., Franks B. W. & Myers T. F. (eds.) (1989). *Papers from the Edinburgh Round Table on the Mental Lexicon.* Edinburgh Working Papers in Cognitive Science, Volume 4. Centre for Cognitive Science, University of Edinburgh. January 1990.

Franks B. W. & Braisby N. R. (forthcoming). Sense Generation or How to Make a Mental Lexicon Flexible. Paper to be presented to the 12^{th} Annual Conference of the Cognitive Science Society, MIT, July 1990.

Myers T. F., Franks B. W. & Braisby N. R. (1989). Partiality and Coherence in Concept Combination. Forthcoming in the Proceedings of the International Colloquium on Cognitive Science, San Sebastian, Spain. May 1989.

Situating Word Meaning

Nick Braisby*

1 Introduction

In psychology, prototype theory stands as a robust and popular statement of what a concept is. But, more recently, prototype theory has also become associated with the issue of word meaning. In particular, it is assumed that the problem of giving some compositional semantics to complex phrases such as "window box" reduces to the problem of how the constituent *concepts* combine. That is, the meaning of a word is described by the concept the word labels. This paper is an attempt to explore the psychological literature as it relates to the problem of describing word meanings. Though we will not be committed to the view that word meanings are concepts, the literature often will be. The next section of this paper outlines a number of puzzles that we would expect a theory of word meaning to solve. The third section contains a brief discussion of the two major psychological approaches to word meaning and a number of morals suggested by this discussion. The fourth section details a Relational View of word meaning which, I shall argue, not only respects these morals but offers solutions to the puzzles described in section two. These solutions are presented in section five and the following section comprises an illustration of how this Relational View may offer a solution to a particular problem in defining kinship relations raised by Betsy Macken (1990). In the last main section, I shall begin to draw out some of the implications of this view. An important step will be to indicate how the Relational View, being psychologically motivated, relates to prototype theory. I will also suggest that Wittgenstein's notion of "family resemblance", which has often been taken as motivation for prototype theory, is equally well, if not better, accounted for by the Relational View. Finally, I will argue that the Relational View is better placed than prototype theory to sit with recent work on concepts, work emphasising their illusory and transient nature. In concluding, I will also suggest that within the Relational View a clear and principled distinction between concepts and word meaning can be drawn. While prototype theory may be a theory of concepts, I shall argue that it cannot be a theory of word meaning.

2 The Puzzles

Each of these puzzles can be thought of as illustrations of the more general problems of defeasibility and polysemy. The first illustrates the simple fact that the application of a word to an object is highly dependent on the situation in which the word is used: in some situations the word appears to apply, in others it appears not to so apply. The third puzzle illustrates the fact that words may, in certain situations, apply to all manner of objects to which they would not ordinarily apply. And the second puzzle illustrates the problem of isolating necessary and sufficient conditions underlying even the conventional applications of words to objects: for example, the application of the word

*The research reported here was supported by an ESRC studentship, award No. C00428722001. My greatest thanks go to Robin Cooper without whom this paper would not have been possible. Many thanks also go to John Perry for his help in suggesting so many improvements to earlier versions of the paper. In addition, I would like to thank Jon Barwise, Jo Calder, Nick Chater, Richard Cooper, George Dunbar, John Etchemendy, Bradley Franks, Liz Hutchinson, Ewan Klein, Karen Lyon, Mark Moens, Terry Myers, Scott McGlashan, Mike Oaksford, Jon Oberlander, Mukesh Patel, Jerry Seligman, Keith Stenning and Max Volino. All errors, of course, are entirely my own.

"lemon" to objects which are undeniably lemons. The second puzzle is really the only one to have been addressed by the two major psychological approaches to word meaning, classical theory and prototype theory. I will argue that all three puzzles can be assimilated to the same framework, the Relational View to be developed later.

2.1 The Lion Puzzle

Consider the following puzzle:

> Fred is sitting on a park bench in London. He knows that at the other end of the park there is a statue of a lion. A schoolgirl approaches him and, explaining that she has been given an assignment to sketch a lion, she asks Fred if he has seen one. Fred replies that he has and points her towards the statue. A little later an exhausted zoo-keeper appears and, explaining to Fred that a lion has escaped from the zoo, he also asks Fred if he has seen one. Fred replies that he hasn't.

How are we to characterise Fred's responses? We might be tempted to say that Fred is being flexible with the truth but such a characterisation seems inappropriate: characterising one response as true, the other as false, would not help us in capturing the fact that both seem equally felicitous. An alternative might be to claim both utterances to be veridical but that "lion" is simply ambiguous: that is, Fred's utterances employ different meanings of "lion", one meaning for real lions, another for statues of lions. Quine (1960), for example, suggests that "light" is ambiguous in just this way. He claims that one "clear condition" of ambiguity is the fact that the application of an ambiguous word to certain objects can be both affirmed and denied: such terms can be "clearly true or clearly false of one and the same thing" (§27 p. 131). A raven's feather, for example, is both light (in weight) and not light (in colour). Similarly, a particular financial institution may be both a bank (a money bank) and not a bank (a river bank). Fred's case seems similar, but whereas we may be led to claim, intuitively, that Fred's statue both is, and is not, a lion, there is an important difference in the cases of "light" and "bank". In these latter cases, the different meanings of these words are deemed independent and unrelated (at least, synchronically). In the case of Fred's uses of "lion", the two meanings are highly related and indeed it is crucial to the success of his utterances that this is so.

In a similar vein, it seems, we wouldn't be tempted to claim that "lion" is simply vague as "heap" and "bald" are considered vague. It is not that the application of the word "lion" to the statue is fuzzy; rather, there is a clear sense in which we can say of the statue that it is a lion and a clear sense in which we can say that it is not. Broadly speaking, it seems then that there are two sorts of meaning attached to "lion": a *core* meaning, to do with *real* lions, and a *peripheral* meaning, to do with other kinds of lion, which is *related* to the core meaning in a particular way. Which meaning the word carries on a particular occasion of use seems to depend on the informational requirements of the agents involved which, in turn, seem to depend on the sort of situation they find themselves in. An adequate theory of word meaning must explain the relationship between core and peripheral word meaning while doing so in a way that supports the basic intuition that the application of a word to an object crucially depends on the nature of the situation in which the word is used.

2.2 Putnam's Lemon

In trying to represent certain aspects of the meaning of words, it seems we want to capture certain generalisations about the properties of the objects to which those words apply. Though puzzles like the previous one may convince us of the difficulty of stating whether a given word applies to a given object, we may be tempted to treat such puzzles as special cases. Accordingly, we may consider the objective of a theory of word meaning to be the identification of those conditions which allow a word to apply to, and only to, objects which it undeniably describes. We might,

for example, consider the conditions which allow the word "lemon" to apply to objects which most certainly are lemons and not to objects (like statues) which most certainly are not. Such a restriction is not one we want to impose, as is made clear below. However, even with such a restriction, such conditions are difficult to isolate, as the arguments of Putnam (1975) testify: though we might want to say that lemons are oval-shape, we can easily imagine discovering one that has been squashed. Similarly, we can imagine finding sweet-tasting lemons, lemons painted red and so on. Virtually every property one might want to ascribe to the class of objects that are lemons can seemingly be defeated. That is, for virtually every property, we can imagine counterfactual circumstances under which a lemon does not possess that property. The problem then is how a theory of word meaning can capture generalisations concerning the properties of a class of objects *and* allow for defeasibility: how is it that we can infer that an object described as a lemon is oval and acidic-tasting and yet accept that some lemons can be flat and sweet?

2.3 Nunberg's Ham Sandwich

Nunberg (1978) discusses a particular case of metonymy (see also Lakoff, 1987): a waiter employs the description "ham sandwich" to refer to a customer who has ordered a ham sandwich. This puzzle is similar to the Lion Puzzle though it offers a starker illustration of the fact that words may apply to all manner of objects to which they would not ordinarily apply. In this case "ham sandwich" applies to an individual (a customer) who has none of the properties of ham sandwiches which make them ham sandwiches.

Similar considerations apply in this case as in the Lion Puzzle. Claiming that "ham sandwich" is ambiguous ignores the important relations between its various uses. Additionally, since we can imagine numerous similar examples, to analyse these in terms of ambiguity would suppose a quite vast, and probably implausibly vast, lexicon. Indeed, this is very much the point which Clark (1983) suggests undermines traditional views of parsing. Similarly, just as before, claiming that "ham sandwich" is vague or questioning the veridicality of the waiter's utterances also seem implausible.

It is also clear that in this case it is the situation in which the waiter finds himself that allows him to make the particular use of "ham sandwich" that he does: if he said the same thing in describing a total stranger in a launderette, it is unlikely he would convey the same meaning. What we require of a theory of word meaning is that it allows the sort of defeasibility illustrated by all these examples to be very much dependent on the nature of the situation.

2.4 Some General Remarks

One of our objectives in this paper is to assimilate puzzles like the above to the same framework. This is something that the Relational View developed later is intended to do. However, one might, at this point, be concerned that the puzzles reveal important differences as well as similarities and that there should be differences in the treatments reflecting these differences in the puzzles. Perhaps the most important difference between these puzzles, it is argued, is that puzzles 1 and 3 are cases where a word applies to objects to which it does not conventionally apply (a stone lion is not really a lion), while puzzle 3 is concerned with the cases of words applying to objects to which they undeniably apply. Even flattened, unripe lemons which have been painted blue are, it may be argued, undeniably lemons. Our position will require focussing on this word "undeniably", but before we can do this we must distinguish two concerns, one semantic, the other ontological, which, it seems, are often conflated.

What we require our theory of word meaning to tell us is under what circumstances a given word may apply to a given object. Indeed, to tell us also to which objects a given word may ordinarily apply. That is, we are concerned with whether a given object can be said to be a "such-and-such". This, then, is the semantic question which we seek to address. Now, this is very different from a

related concern which is with whether something actually *is* a such-and-such. This, then, is the ontological question for which we do not intend seeking an answer. To exemplify, in the case of, say, a stone statue of a lion, we require our theory to outline some of the conditions of word use which allow the word "lion" to apply to such an object. One can, after all, it seems, say of such an object that it is a "lion". But this use of the word does not commit one, and nor should it commit our theory, to the view that such an object *is* a lion. Correspondingly, it is precisely those conditions under which something is or is not a lion, say, which we do not require our theory to outline. Instead, we require the theory to outline those conditions under which words, such as "lion", can be applied to various objects. It is important to see the distinction between these concerns. The fact that under traditional theories of meaning, as exemplified by the arguments of Kripke (1972) and Putnam (1975), the two concerns are run together forms the subject of another discussion (Braisby, 1990). These concerns, the ontological and the semantic, can be separated, however, and it is on this fact that our approach relies.

So, returning to the sense in which words can be said to apply "undeniably" to certain objects, we can now say this. First, that natural science may well inform us that squashed lemons really are lemons, just as it may, conceivably, inform us that they are not. In this sense then, squashed lemons may undeniably be lemons. Second, that this is not necessarily what should concern a theory of word meaning, though it does according to traditional theories. Rather, we can sum up our concern as follows. Our theory should tell us under what circumstances an object can be said to be a "such-and-such". It is for natural science to tell us whether the object really is a such-and-such. The relation between the fruits of natural science and the question as to which objects a given word may apply is, presumably, complex and any account of it involved. Such an account is beyond the scope of this investigation.

Setting aside any ontological concerns and concentrating on the semantic does not, it could be said, necessarily lead to our treating the puzzles in the same way. It might be suggested, for example, that whereas it *can* be said of a stone lion that it is not a lion, it *cannot* be said of a squashed lemon that it is not a lemon. And this regardless of what natural science may tell us. After all, it may be true that natural science is such as to convince us that squashed lemons really are lemons and that stone lions really are not lions, but, as we have suggested, this is a strictly separate issue from that of the word "lemon" applying or not applying to various objects. Rather, to show the semantical similarity between puzzle 2 and the others, we need to show that there are circumstances under which the word "lemon" would not apply to certain entities which, according to natural science, are nonetheless lemons. Consider the following example. An assistant, working at the local supermarket, is asked to put onto the display stand all the lemons to be found in the warehouse. In the execution of this duty, the assistant comes across many lemons, which nonetheless she doesn't return to the stand: lemons which are diseased, squashed, unripe and so on. No doubt there are alternative analyses, but the one we pursue here is to assume that this use of lemon, in this context, is such as to apply to some lemons and not others. That is, it applies to normal, saleable lemons but not unusual, unsaleable ones. Other examples could be constructed. Consider the issue of whether "lemon" applies to a lemon which has just been liquidised. It seems, that here too, one can imagine cases where we would say that the word "lemon" would not apply to such a mass. When, for example, someone is asking for a lemon in order to slice it and drop it into a gin and tonic. Now, such uses may not be as common as those in the other puzzles and in this, the puzzles differ. But in the fact that such uses are possible lies the similarity between all the puzzles. And it is this similarity which leads us to expect them to be assimilated to one and the same framework.

3 The Morals

In this paper our concern is with a psychologically plausible account of word meaning. Psychological approaches to word meaning have become associated with the view that, in some way,

the meaning of a word can be described by the concept the word labels. For example, Cohen & Murphy (1984) describe the task of arriving at a compositional semantics for phrases such as "ocean drive" as involving finding "a mediating relation between *concepts*"[1]. In this section, we will briefly examine the two major psychological approaches to concepts, classical theory (see, for example, Hull, 1920; Vygotsky, 1962) and prototype theory (Rosch, 1973, 1975; Rosch & Mervis, 1975) before discussing what they tell us about word meaning.

If we take Putnam's lemon as an example, we can consider the sort of information that these theories might have us encode in representing the concept "lemon". We can describe these concepts in terms of attribute-value matrices that are at least implicit (and sometimes explicit) in the discussions of many researchers in the area (for example, Murphy & Medin, 1985; Osherson & Smith, 1981; Smith & Osherson, 1984; Smith et. al., 1988).

A Classical description of the Concept "lemon"

$$\begin{bmatrix} \text{shape:oval} \\ \text{colour:yellow} \\ \text{taste:acidic} \\ \text{isa:fruit} \end{bmatrix} \tag{1}$$

A Prototype description of the Concept "lemon"

$$\begin{bmatrix} \text{shape:} \left\{ \begin{matrix} \text{oval:9*} \\ \text{square:1} \end{matrix} \right\} \\ \text{colour:} \left\{ \begin{matrix} \text{yellow:8*} \\ \text{green:2} \end{matrix} \right\} \\ \text{taste:} \left\{ \begin{matrix} \text{acidic:8*} \\ \text{sweet:1} \end{matrix} \right\} \\ \text{isa:fruit:9*} \\ \vdots \end{bmatrix} \tag{2}$$

In (1) we have a simple attribute-value structure, which we can view as expressing a conditional constraint on inferences that the concept may licence: if we have an object X and X has acidic taste, yellow colour and so on, then we can infer that X is a lemon[2]. Alternatively, and especially from a consideration of word meaning, we can view (1) as a kind of "definition" or "lexical decomposition" and the corresponding conditional as a "meaning postulate" (Cresswell, 1978; also Fodor et. al., 1980). In (2) the attribute-value structure has been augmented in a fashion similar to Smith & Osherson (Smith & Osherson, 1984; p.360[3]). Paths within {} braces are disjoint. The numbers appearing after the values represent a weighting on the disjunction, higher numbers indicating greater typicality. That is, a lemon with acidic taste is considered a more typical lemon than one with sweet taste. The starred paths indicate the *prototypical* attribute-value pairs, those possessed by the prototype for lemon[4]. This representation not only allows the possibility of a lemon having a square shape and a sweet taste, but it also indicates that lemons having oval shape and acidic

[1] The emphasis is my own.

[2] This expresses the fact that such conditions can be seen as sufficient. Alternatively we can view the attribute-value structure as expressing the fact that such conditions may be seen as necessary. In that case the conditional would be reversed: if X is a lemon then we can infer that X has acidic taste, yellow colour, etc.

[3] In fact, this picture is a simplified version of what Smith & Osherson describe. In their 1984 paper, Smith & Osherson also talk of additional weightings they term diagnosticities. This omission is made solely for the purposes of clarity and does not affect the arguments made later.

[4] The prototype is a theoretical construct that may or may not correspond to an actual exemplar.

taste are more typical. Having briefly considered how classical and prototype theories describe concepts (and how, it is assumed, they describe word meanings) we are now in a position to explore in a little more detail some of the problems associated with these theories. In doing this I shall draw a number of morals that we would expect any psychologically plausible account of word meaning to respect.

Moral 1 *The Moral of Coherence*

Categories cohere. That is, there is *something* that brings objects together to form categories and whatever concepts are it is this "something" that they should specify. For classical theory, coherence is determined by attribute-value matching. That is, to determine whether a given object is a member of a category, we must match the attribute-value pairs of the two attribute-value structures corresponding to that object and that category. For prototype theory too, coherence is determined by attribute-value matching. There are a number of arguments, though, why an attribute-value matching view of coherence is going to be insufficient to capture the full rich structure of real-world categories.

Recent discussion has led to the view that representing coherence is necessarily going to involve what might be described as "knowledge-rich" techniques. As a simple example we might consider how a squashed, sweet lemon that has been painted blue is still categorised as a lemon. Seemingly, this can't be modelled in terms of similarity as computed by attribute-value matching since very few of the attribute-value pairs of either (1) or (2) would match those describing the object. Murphy & Medin (1985) and Medin & Wattenmaker (1987) argue that similarity computed by attribute-value matching is simply insufficient to determine the coherence of a category. Instead, Murphy & Medin (1985) propose that a major part of the coherence of a category is determined by "theories and knowledge of the real world" (p.312), these, in turn, being composed or "made up" (p.313) from concepts. Theories like those, for example, which express knowledge of such relations as flattening, painting, ripening, etc.

In trying to glean psychological truths for our account of word meaning this position seems circular. If we accept for sake of argument that word meanings can be described by concepts then Murphy & Medin's arguments suggest that word meanings are to be represented in terms of theories which in turn are to be represented in terms of word meanings. Though, Murphy & Medin reject the suggestion that their position is circular, a clearer rejection can be made, I believe, once a particular assumption is abandoned: the Unitary assumption.

Underlying both the prototype and classical approaches is an assumption, the Unitary assumption, that a single, "basic-level" category (Rosch et. al., 1976) is to be represented by something corresponding to a single, unitary data structure, variously called a concept, lexical concept or word meaning. It seems that it is this assumption together with Murphy & Medin's arguments concerning coherence which appear to lead to circularity. The problem arises because a word meaning is seen as both being a component of a theory and being composed of theories. For an account of word meaning, this suggests that underlying various uses of a single word are various word meanings (that is, as components of theories). The circularity emerges when we stipulate that, in fact, it is only one meaning that underlies all the uses of a given word. Thus we are lead to represent meanings inside meanings and the circularity that implies. We can avoid this circularity by denying that the various uses of a word all depend on a single meaning. Rather, we may allow that these various uses may depend on a number of different, though related, meanings. To conclude, any theory which attempts to represent all the meanings of the uses of a word in terms of a single data structure can be ruled out by the moral of coherence *a priori*.

As an alternative, it seems that a much more intuitive account of conceptual coherence is forthcoming when we consider some more examples like the lemon. Intuitively, it seems that a dinner-table turkey is a turkey, not because of some *particular* set of properties that it has, but because we

know of relations that hold in the world such that the properties of the object we observe (dinner-table turkey) stand in this relation with the properties of normal turkeys. Similarly, we can still talk of a dismantled bicycle as a bicycle apparently not because of some similarity as computed by attribute-value matching but because of our knowledge of dismantling operations. Murphy & Medin (1985) make a similar point. They state that "much of our reasoning about concepts may be based on constraints about operations that are permissible" (p.295). This point seems fundamental and we might expect that this clue could help in developing our account of word meaning. In particular, if some of the data structures which underly the uses of a word, express facts about such permissible "operations", then it seems our account will have more promise in respecting the psychological arguments regarding coherence.

Moral 2 *The Moral of Representational Economy*

In principle, there seem to be many ways of representing a category by encoding information in our data structures. One way might be to record, for every exemplar, the value of each attribute pertaining to the category. For example, if shape is relevant to the category of lemons then for each lemon we encounter, we could simply record its shape. This is the strategy that prototype theory adopts: if, equipped with (2) as our representation of the category of lemons, we encounter a flat lemon, the prototype structure will be revised. The disjoint set of values for the attribute "shape" will be expanded to include "flat" with a suitable weighting indicating its typicality. Such a strategy seems grossly inefficient and also fails to respect the following intuition: concepts are hypothesised to capture generalisations in order that finite brains may represent infinite possibilities. The strategy also seems implausible from the point of view of cognitive development, a point that has been made by others (for example, Cohen, 1983; p.87). As an alternative, we might consider that it is our knowledge of operations and other relations in the world that allow us to represent categories in the way that we do. Such knowledge should not be encoded in each representation we have for different classes of objects as it is in prototype theory: we wouldn't want to add the value "flat" for the attribute "shape" for each of our representations for "lemon", "orange", "house", "pie" and so on. Rather we would like to say that our knowledge of the "flattening" operation is generalised across our representations of categories. This notion of generalisation of knowledge of operations will be an important feature of the framework to be developed. The moral of representational economy is that we must require that whatever structures we hypothesise to describe word meanings they must genuinely generalise; in principle, they shouldn't incorporate information specific to each exemplar.

Moral 3 *The Moral of Central Exemplars*

In everyday reasoning the inferences we make regarding categories are often based on typical exemplars. If we have to explain what dogs are we tend to describe what have been called their default properties. Such an observation is made much use of by Putnam (1975) and these default properties seem to be very similar to what he calls "core facts". The pervasive nature of default reasoning suggests that the central exemplars, or those exemplars for which "core facts" are indeed facts, have special cognitive significance. This is also something that a theory of word meaning should respect. Classical theory singularly failed to respect this moral and we might see this as the reason for its problems. Prototype theory, on the other hand, accords the central exemplars a special status in that they are represented by the prototype. Due to the weightings on attribute-value pairs, prototype theory also offers a way of distinguishing different, non-central exemplars. However, the moral of Central Exemplars is that any theory of word meaning must distinguish between the representations of central and peripheral exemplars.

Moral 4 *The Moral of Context Dependence*

In a number of ways the meanings of words can be said to depend on the context in which we find them. "Lion" when used to describe a fierce lion seems to mean something quite different from "lion" when used to refer to a statue of a lion. Typically, it appears as though the same word carries

different meanings and the nature of the difference directly reflects the nature of the differences between the situations in which it is used and the purposes for which it is used. Barsalou, for example, has shown very convincingly a number of effects of context on our representations of categories (Barsalou, 1982; Barsalou, 1987). Barsalou's conclusion seems to imply that there are no stable representations corresponding to concepts (a conclusion that the framework we develop in the next section will endorse). However, were the same conclusion to apply to word meanings, it would appear very unintuitive: words do seem to have a stable core meaning and the meaning of words in use seems related to the core meaning. The two meanings of "lion" in the lion puzzle, for example, seem highly related. Therefore the conclusion I will take from observations of context dependence is that a theory of word meaning must have an adequate account of the role of situational factors in determining the deviation of the meaning of words in use from what I take to be their core meaning. This is the Moral of Context Dependence.

4 A Relational View of Word Meaning

In the previous section, we considered various arguments which suggest that both classical and prototype theories run into problems in describing concepts. From the perspective of these psychological theories, similar problems are going to face their accounts of word meaning. However, it seems that if we follow the morals in developing an account of word meaning, we might not only develop a psychologically plausible account but one that may also solve the puzzles described in section two.

The morals allow one a considerable degree of freedom and the framework I shall present here should by no means be seen as the only approach consistent with them. As far as an account of word meaning is concerned, these are the morals we must respect:

1. Descriptions of word meanings must express generalisations and lend themselves to economical representations *a la* classical theory. In addition, we would expect that facts about operations (such as flattening, drying, ripening, etc.) should be similarly expressed in an economical way.

2. Multiple word meanings (perhaps especially those corresponding to operations) should be implicated in the various uses of a single word.

3. Descriptions of word meanings must be sensitive to the important cognitive distinction between central and peripheral exemplars. That is, we expect there to be some distinction between the meanings a word has when used to describe central as opposed to peripheral exemplars.

4. The deviation of peripheral word meaning from core meaning should be sensitive to the situation in which the words are used and to the informational requirements of the agents involved in their use.

In describing the framework I shall offer an interpretation of it in situation theory. Though I shall not refer to it directly, an implementation in C-prolog has been completed.

4.1 Worms

We take it that a word is related to a data structure which describes that word's meaning. We will label such a relation, WORd Meaning (or WORM). Taking the example of lemons once more, the WORM will relate the word "lemon" to a data structure just as in (1) which is repeated below. This Relational View does not depend on the particular attribute-value pairs chosen in (1), that is, it does not depend on whether (1) does indeed describe the core meaning of "lemon". Rather, what the Relational View amounts to is a more general claim about the nature of such core meanings, their relation to peripheral meanings and how language users may be flexible in their use of words.

The data structure for "lemon"

$$\begin{bmatrix} \text{shape:oval} \\ \text{colour:yellow} \\ \text{taste:acidic} \\ \text{isa:fruit} \end{bmatrix} \tag{3}$$

In Situation Theory we can view this WORM as a relation in the following infon where the data structure is represented by a property.

$$\ll \text{WORM}, \text{LEMON}, \text{P}; 1 \gg \tag{4}$$

where

$$\begin{aligned} \text{P} = \quad [\dot{\text{p}} \mid \; & \ll \text{shape}, \dot{\text{p}}, \text{oval}; 1 \gg \wedge \\ & \ll \text{taste}, \dot{\text{p}}, \text{acidic}; 1 \gg \wedge \\ & \ll \text{colour}, \dot{\text{p}}, \text{yellow}; 1 \gg \wedge \\ & \ll \text{isa}, \dot{\text{p}}, \text{fruit}; 1 \gg] \end{aligned} \tag{5}$$

As suggested earlier, WORMs[5] can be thought of as expressing constraints. One constraint might be used in parsing so as to constrain the possible interpretations associated with a given use of a word. For example, given a use of "lemon" we may want to conclude that we have an individual which has acidic taste, yellow colour and so on. Another constraint might be used in generation so as to constrain the possible ways of describing a given individual. For example, given an individual which has the properties of acidic taste, yellow colour and so on, we would wish to conclude that the word lemon may be used to describe such an individual. That is, there are two possible constraints, one which necessarily seems to involve word use, the other which involves the abstract properties of a word. Here I shall be concerned solely with the latter kind of constraint.

Our WORM will associate with a word a particular data structure which expresses the sort of constraint we need to characterise the conditions under which the word can be used. Taking the example of lemons again and assuming that we can also associate a data structure with an individual, then provided that the data structure associated with the individual is subsumed by that associated with the word then the individual can be described as a lemon. In Situation Theory, we will represent the data structure associated with the individual by means of a restricted parameter. What we then have amounts to the following: provided the restricted parameter associated with the object has the property given by the WORM then that word can be used in describing the object. Following Barwise (1985), the constraint can be expressed as follows.

$$\text{S}_1 \Rightarrow \text{S}_2 \mid \text{B} \tag{6}$$

where

$$\text{S}_1 = [\dot{\text{s}} \mid \dot{\text{s}} \models \ll \text{P}, \dot{\text{p}}; 1 \gg] \tag{7}$$

and

$$\text{S}_2 = [\dot{\text{s}} \mid \dot{\text{s}} \models \ll \text{DESCRIBE}, \dot{\text{p}}, \text{LEMON}; 1 \gg] \tag{8}$$

and

$$\text{B} = [\dot{\text{b}} \mid \dot{\text{b}} \models \ll \text{WORM}, \text{LEMON}, \text{P}; 1 \gg] \tag{9}$$

and

$$\begin{aligned} \text{P} = [\dot{\text{p}}_1 \mid \; & \ll \text{shape}, \dot{\text{p}}_1, \text{oval}; 1 \gg \wedge \\ & \ll \text{taste}, \dot{\text{p}}_1, \text{acidic}; 1 \gg \wedge \\ & \ll \text{colour}, \dot{\text{p}}_1, \text{yellow}; 1 \gg \wedge \\ & \ll \text{isa}, \dot{\text{p}}_1, \text{fruit}; 1 \gg] \end{aligned} \tag{10}$$

[5]More correctly, we mean the data structure which is an argument of the WORM. Throughout the following, however, we will use WORM to mean both this data structure and the relation. Although the distinction is not unimportant, which is meant will be clear from the text.

(6) is a conditional constraint. That is, it is a constraint that is relativised with respect to some background situation type, B, such that provided we have a background situation of this type then the constraint holds. (6) and (9) express the fact that if we have an individual with the property P then it is in virtue of the WORM "lemon", a component of the background conditions, that the individual can be described as a lemon[6].

So far we have a framework that is very similar to classical theory and we might consider whether it respects the morals that are so important. Although we cannot yet see how all the morals are to be respected, we can note in passing that the moral of representational economy is respected since our data structures capture a significant degree of generalisation, *a la* classical theory. WORMs also offer a way of describing the central exemplars. The question then remains as to how to accommodate the peripheral exemplars and, as we shall see, these are to be accommodated by what we will call COMBINATIONS of WORMs (or COWORMs).

4.2 Coworms

Consider a peripheral exemplar of the category of lemons, one that has been squashed flat. With this individual we can also associate what in Situation Theory can be modelled by a restricted parameter. The parameter this time though will not have the complex property given by the WORM for lemon but, instead, will be as follows.

$$\begin{aligned} \dot{p} \mid\ & \ll \text{shape}, \dot{p}, \text{flat}; 1 \gg \wedge \\ & \ll \text{taste}, \dot{p}, \text{acidic}; 1 \gg \wedge \\ & \ll \text{isa}, \dot{p}, \text{fruit}; 1 \gg \wedge \\ & \ll \text{colour}, \dot{p}, \text{yellow}; 1 \gg \end{aligned} \tag{11}$$

Since $\dot{p}$ is not of the type expressed by the WORM for lemon, the constraint (6) cannot allow that this object can be described by a use of the word lemon. However, cognitive agents have a wealth of information at their disposal that we might characterise as being in their background resource situation. This situation allows the agent access to information about the world that will resolve certain local inconsistencies. One such inconsistency is where the object we are considering does not have the properties expressed by the WORM for lemon (e.g., a squashed lemon) and yet may nonetheless be described as a lemon. In this instance the resource situation allows us to access WORMs that can resolve the inconsistency. One such WORM is that for "flattened". This WORM is also a relation between a word, "flattened", and two sorts of data structures. One data structure will describe flat individuals, the other will describe individuals of indeterminate (though not null) shape.

The arguments we considered in discussing coherence and representational economy indicate the kind of solution we require. We need a further constraint which expresses the fact that an individual with a certain set of properties may be described as, say, a lemon, provided that the properties it has are related, by a WORM, to those of the WORM "lemon". So we will have a WORM "lemon" and a WORM "flattened" which when combined suitably will describe a flattened lemon. In Situation Theory, this can be expressed as another constraint governing the properties of the word "lemon". Provided the restricted parameter associated with a given object is of a sort that is related by the WORM for flattened to the sort of parameter associated with the WORM for lemon, then the object is such that it could be described by a use of the word lemon. This new constraint is as follows.

$$S_1 \Rightarrow S_2 \mid B' \tag{12}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \tag{13}$$

[6] Although I shall not discuss the issue of perspectives, by locating WORMs in the background situation type this account seems to have certain interesting parallels with the notion of perspectives developed by Jerry Seligman (1990).

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{LEMON}; 1 \gg] \tag{14}$$

and

$$B' = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LEMON}, P_1; 1 \gg \wedge \ll \text{WORM}, \text{FLATTENED}, P_1, P_2; 1 \gg] \tag{15}$$

and

$$P_2 = [\dot{p} \mid \ll \text{shape}, \dot{p}, \text{flat}; 1 \gg \wedge \ll \text{taste}, \dot{p}, \text{acidic}; 1 \gg \wedge \ll \text{colour}, \dot{p}, \text{yellow}; 1 \gg \wedge \ll \text{isa}, \dot{p}, \text{fruit}; 1 \gg] \tag{16}$$

Very briefly, we have seen how this Relational View can cope with both central and peripheral exemplars. Respecting the moral of central exemplars, it treats central and peripheral exemplars differently: central exemplars are accommodated by WORMs, peripheral exemplars by COWORMs. The fact that peripheral exemplars are treated by COWORMs, however, allows the moral of coherence to be respected. For instance, in the treatment we gave for our flattened lemon, one can see the WORM "flattened" as expressing a theory. It is precisely this kind of "knowledge-rich" solution that the moral of coherence demanded. The moral of context dependence is also respected by ensuring that the WORMs that underly word use are seen as components of the background situation. The background situation may change either as agents focus situations change or as agents require different information from their focus situation. In addition, COWORMs capture the added generalisation that prototype theory seems to miss. Our knowledge of flattening is not encoded in each of our representations for various categories but it is encoded separately in a way that will generalise. This plausible and desirable result suggests an interesting developmental prediction. Namely, when we acquire the WORM "flattening" we concomitantly learn that all manner of objects that are flat can be described in new ways: we learn that flat objects that otherwise look like lemons can also be described as lemons. Unlike most other accounts, this allows that we don't have to change our data structure describing lemons on encountering a flat lemon for the very first time.

In addition to respecting the Morals, I believe the account offered here is promising for other reasons. Consider a situation in which Fred comes into my office and tells me that he has a lemon in his pocket. Corresponding to an instance of default reasoning, there is a sense in which I come to believe that Fred has a yellow, oval-shaped fruit on his person. This inference can be expressed by considering that it is the WORM for lemon which is used in such reasoning. We may, however, acquire further information that conflicts with the WORM for lemon and so rules out "lemon" as an accurate description. In such circumstances we may have to find a different appropriate COWORM or COWORMs. However, there is also a sense in which I know nothing of what Fred has in his pocket other than that there is something and Fred calls that something a lemon. The fact that so little information can be conveyed by the use of a word seems to correspond exactly to the indeterminacy in finding a particular corresponding WORM or COWORM to describe that word's meaning.

COWORMs are also implicated in the realisation of the background conditions with respect to which conditional constraints are relativised. When we wished to describe a squashed lemon as a lemon, the first constraint (6) clearly failed as (11) does not have the property P in (9). The inapplicability of the constraint we can assume to be due to the fact that the actual background situation was not of type B in (6). The inappropriateness of the constraint leads to the utilisation of a different but related constraint as in (12), relativised with respect to some new set of background conditions, B'. This time the constraint holds because (11) has the property P_2 in (15). If we consider the way this constraint is expressed then we can see it as revealing something of the nature of the original background conditions with respect to which the constraint in (6) is relativised. The particular COWORM that is part of the second constraint tells us that the negation of that COWORM was

a condition of the original background situation type. That is, returning to our example, the fact that the WORM "flattened" relates the constraint (12) to the constraint (6) tells us that (6) presupposes that lemons aren't squashed flat and this then is one of the background conditions with respect to which (6) was relativised. In this framework then, the process of finding an appropriate COWORM is the process of realising background conditions.

So now we have a very basic framework with which to explore word meaning. It seems to respect the morals that I have tried to argue are so important and by the use of COWORMs the account seems to offer a solution to the problem of Putnam's Lemon (puzzle 2): one can imagine similar analyses to that offered for some of Putnam's other examples. Lemons, having turned blue in the presence of some gas, for example, might be accommodated by some WORM or COWORM expressing the relation between affected and unaffected lemons. But what of the other puzzles in section 2? Can the framework begin to explain what is going on? I believe that it can, in a most natural and appealing way.

5 The Puzzles Reconsidered

Fred has been sitting on a park bench experiencing a number of situations. In one situation he claims to have seen a lion and in this we would probably agree. In another, he claims not to have seen a lion and again we wouldn't argue. Now we can suppose that Fred knows that something like the following constraint holds concerning what kind of object may be described as a lion.

$$S_1 \Rightarrow S_2 \mid B \tag{17}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P, \dot{p}; 1 \gg] \tag{18}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{LION}; 1 \gg] \tag{19}$$

and

$$B = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LION}, \dot{P}; 1 \gg] \tag{20}$$

and[7]

$$P = [\dot{p} \mid \ll \text{animate}, \dot{p}; 1 \gg \wedge \ll \text{legs}, \dot{p}, 4; 1 \gg \wedge \ll \text{tail}, \dot{p}, 1; 1 \gg] \tag{21}$$

When Fred responds to the zoo-keeper, his use is in accordance with this constraint. That is, Fred *hasn't* come across an object whose description is of the sort expressed by the WORM for lion, i.e., an animate object with 4 legs and a tail. So provided that the constraint Fred is exploiting is expressed by the WORM "lion", he has replied correctly to the zoo-keeper. But what of the schoolgirl? Fred has access to a background resource situation which supports many infons. One expresses the fact that the there is a WORM "statue". This WORM relates two data structures to the word "statue". One data structure may describe animate things (the class of things modelled by statues), the other inanimate things that stand in certain relations to the animate things (the class of statues). Though this doesn't accurately or fully describe the relationship between statues and the things they model, it will suffice for our purposes. In a similar fashion to before, the WORM "statue" allows us to arrive at the following constraint.

$$S_1 \Rightarrow S_2 \mid B' \tag{22}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \tag{23}$$

[7] Clearly, P only expresses *some* of the information relevant to lionhood.

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{LION}; 1 \gg] \tag{24}$$

and

$$B' = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LION}, P_1; 1 \gg \wedge \ll \text{WORM}, \text{STATUE}, P_1, P_2; 1 \gg] \tag{25}$$

and

$$P_2 = [\dot{p} \mid \ll \text{animate}, \dot{p}; 0 \gg \wedge \ll \text{legs}, \dot{p}, 4; 1 \gg \wedge \ll \text{tail}, \dot{p}, 1; 1 \gg] \tag{26}$$

Now Fred, whilst walking to his park bench, *has* come across an individual whose description is of this last sort, i.e., an inanimate object with 4 legs and a tail. So, provided that the constraint he is exploiting is expressed by this COWORM, Fred can rightly assert that he has seen a lion which is, fortunately, what the schoolgirl was hoping for.

What should be clear is that COWORMs are sensitive to the information states of agents. If Fred's background resource situation did not support the infon concerning the WORM for "statue" then, according to this view, he would not be able to rightly report that he had seen a lion. The significance of this point is much more obvious when we consider the third puzzle of section 2.

The waiter in the restaurant knows (we can assume) that "ham sandwich" is normally appropriate to describe ham sandwiches and not human beings like the customer he is serving. However, his particular background resource situation also supports the fact that the WORM relation holds between the word "orders", things of type "customer" (P_2) and things of type "ham sandwich" (P_1). Just as before, this WORM can form a COWORM with the WORM for "ham sandwich" which allows us to consider a new background situation type as follows.

$$B' = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{HAM\_SANDWICH}, P_1; 1 \gg \wedge \ll \text{WORM}, \text{ORDERS}, P_2, P_1; 1 \gg] \tag{27}$$

Similar to Fred's case, we now have a COWORM which expresses a constraint relativised with respect to this background situation type. It is this new constraint that renders it legitimate for the waiter to refer to his customer as a ham sandwich. The point about this particular example that is not quite so clear in the others I have described is precisely the contextual nature of this aspect of word use. It seems reasonable to suggest that there is something about the particular background resource situation that allows the waiter to call a customer a ham sandwich. Locating such a mechanism in the background situation seems to offer a way of explaining how it would be considered infelicitous were the waiter to describe the same person in the same way in a launderette. It seems the particular COWORM we have considered would not be supported given the nature of those prevailing background conditions.

So far, I have tried to argue that the Relational View of word meaning can offer solutions to the sorts of puzzle exemplified by those of section 2. Now we turn to another puzzle, that provided by kinship terms, and apply the same kind of analysis.

6 Lakoff's Mother and Macken's Father

The problem of defining kinship relations has been made clearer by the accounts of Lakoff (1987) and Macken (1990). Mothers, for example, are conventionally considered to be genetically related to their offspring. Adoptive mothers, though, are naturally described as mothers even though they and their adopted children are not related genetically. Similarly, as Macken's example demonstrates, fathers may be live-in but are fathers nonetheless. The Relational View I have outlined suggests an alternative analysis to those offered by both Lakoff and Macken which we will now

pursue. Under this view, we might like to say that the following constraint holds concerning the sort of objects that can be described by a use of the word "mother".

$$S_1 \Rightarrow S_2 \mid B \tag{28}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P, \dot{p}; 1 \gg] \tag{29}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{MOTHER}; 1 \gg] \tag{30}$$

and

$$B = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{MOTHER}, P; 1 \gg] \tag{31}$$

and

$$\begin{aligned} P = [\dot{p} \mid\ & \ll \text{sex}, \dot{p}, \text{female}; 1 \gg \wedge \\ & \ll \text{child\_of}, \dot{p}, \dot{q}; 1 \gg \wedge \\ & \ll \text{genetics}, \dot{p}, \dot{g}; 1 \gg \wedge \\ & \ll \text{genetics}, \dot{q}, \dot{g}; 1 \gg] \end{aligned} \tag{32}$$

It is this constraint that tells us that a female who is genetically related to her child, may be described by a use of the word "mother"[8]. This, however, does not help us describe adoptive mothers: we have no mechanism for stating that an individual described by the following restricted parameter can be described by a use of the word "mother".[9]

$$\begin{aligned} \dot{p} \mid\ & \ll \text{sex}, \dot{p}, \text{female}; 1 \gg \wedge \\ & \ll \text{child\_of}, \dot{p}, \dot{q}; 1 \gg \wedge \\ & \ll \text{genetics}, \dot{p}, g_1; 1 \gg \wedge \\ & \ll \text{genetics}, \dot{q}, g_2; 1 \gg \end{aligned} \tag{33}$$

In a similar fashion to before, we will use some other WORM to form a COWORM which provides a further constraint on the way in which the word "mother" can be used. We will assume that agents have access to the WORM for "adoptive". This WORM relates the word "adoptive" and two data structures. One data structure expresses the fact that two individuals have the same genetic material, the other expresses the fact that their genetic material is unrelated[10]. This WORM, combined with the WORM "mother" gives us a COWORM that expresses the following constraint.

$$S_1 \Rightarrow S_2 \mid B' \tag{34}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \tag{35}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{MOTHER}; 1 \gg] \tag{36}$$

and

$$\begin{aligned} B' = [\dot{b} \mid \dot{b} \models\ & \ll \text{WORM}, \text{MOTHER}, P_1; 1 \gg \wedge \\ & \ll \text{WORM}, \text{ADOPTIVE}, P_1, P_2; 1 \gg] \end{aligned} \tag{37}$$

and

$$\begin{aligned} P_2 = [\dot{p} \mid\ & \ll \text{sex}, \dot{p}, \text{female}; 1 \gg \wedge \\ & \ll \text{child\_of}, \dot{p}, \dot{q}; 1 \gg \wedge \\ & \ll \text{genetics}, \dot{p}, \dot{h}; 1 \gg \wedge \\ & \ll \text{genetics}, \dot{q}, \dot{g}; 1 \gg] \end{aligned} \tag{38}$$

[8] As described here, we have said that the child and mother must have the *same* genetics. Though this is most certainly not the correct condition on the use of the word "mother" it will suffice for illustrative purposes.

[9] In (33), g_1 and g_2 are assumed to be distinct.

[10] Again, this is not the correct condition but will suffice for illustrative purposes.

It is this constraint, expressed by the COWORM of our WORMs for "adoptive" and "mother", that allows that individuals who are adoptive mothers and have no genetic relation to their children can nonetheless still be described by a use of the word "mother". Importantly, this analysis does not lead to the suggestion that there are several independent models underlying the word "mother" as Lakoff's analysis does. Nor does it require that we have to accept the apparently unintuitive suggestion of Macken's analysis, that the "father" relation is really a relation between a father, a child and some parameter.

7 Some Implications

In this section we briefly consider three of the implications of this Relational View of word meaning. Just as previous psychological theories of concepts have been advanced as theories of word meaning, so one of our concerns is to consider the implications for a theory of concepts of our view of word meaning. Another concern is with Wittgenstein's "family resemblance". As other psychological theories have claimed to be able to express family resemblance, it is important to examine the capabilities of the Relational View in this regard. Our first concern, however, is more directly psychological. It is to attempt, in some detail, a comparison between the Relational View and prototype theory. The means of achieving this is to effect a *reconstruction* of prototype theory from the Relational View, noting what is lost in the process.

7.1 On Prototype Theory

In this section, we return to prototype theory, more properly, a variant of this known as the *Modification Model* (Smith & Osherson, 1984), in an attempt to provide some better understanding of it and of the Relational View we have developed. As already indicated, we will attempt to *reconstruct* prototype theory out of the Relational View and note the consequences. Accordingly, our goal is to arrive at some rough translation of the sort of data structure we have in (2).

Whereas we are attempting to reconstruct prototype theory by "tinkering" with the Relational View, this is purely for expository purposes. As I have tried to argue, the Relational View receives much stronger support from psychological considerations than does prototype theory. It is via this "tinkering" that we hope to understand better what is wrong with prototype theory and, concomitantly, what is right with it.

The Relational View is tied to the claim that a word's core meaning may be described by a single data structure. This serves to describe that word's meaning on certain occasions of use and acts as one constraint on the word's meaning. For different occasions of use we have seen that different sorts of data structure are implicated. In particular, those involving relations such as "flattening", etc., hence, the Relational View. That is, underlying word use generally is a whole family of related constraints or meaning relations. Our first goal in reconstructing prototype theory is to try and reduce this family of constraints to one. By these means, we will be able to see more clearly how to retrieve the single data structure that prototype theory postulates as underlying word use. The fact that this one data structure is supposed to account for all the meanings of a word, means that, in terms of constraints, we should be trying to reduce our family of conditional constraints not merely to one constraint, but to one unconditional constraint. These then are the moves we have to make. Schematically, our starting position, the Relational View, is as follows, where PROPERTY(P, x) means the Situation Type where x has the property P and DESCRIBE(x, LEM) means the Situation Type where x can be described by a use of the word "lemon". Similarly, WORM(LEM, P), B_1, etc., all refer to Situation Types. Hopefully it is clear that this notation is not intended to be formally precise, but rather to provide a simple and intuitive illustration.

$$\begin{aligned}&\text{PROPERTY}(P, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \mid \text{WORM}(\text{LEM}, P) \wedge B_1 \\ &\text{PROPERTY}(Q, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \mid \text{WORM}(\text{LEM}, P) \wedge \text{WORM}(\text{FLAT}, P, Q) \wedge B_2 \\ &\text{PROPERTY}(R, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \mid \text{WORM}(\text{LEM}, P) \wedge \text{WORM}(\text{PAINT}, P, R) \wedge B_3\end{aligned} \tag{39}$$

That is, (39) expresses schematically the Relational View that we have developed for those individuals which are, respectively, ordinary lemons, flattened lemons and painted lemons. Now, in order to make these conditional constraints unconditional the background conditions must be made explicit. Doing so yields the following family of constraints.

$$\begin{aligned}&\text{PROPERTY}(P, x) \wedge \text{WORM}(\text{LEM}, P) \wedge B_1 \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \\ &\text{PROPERTY}(Q, x) \wedge \text{WORM}(\text{LEM}, P) \wedge \text{WORM}(\text{FLAT}, P, Q) \wedge B_2 \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \\ &\text{PROPERTY}(R, x) \wedge \text{WORM}(\text{LEM}, P) \wedge \text{WORM}(\text{PAINT}, P, R) \wedge B_3 \Rightarrow \text{DESCRIBE}(x, \text{LEM})\end{aligned} \tag{40}$$

If we make the further assumption that B_1-B_3 will always hold[11] and we abstract over the WORMs we get the following.

$$\begin{aligned}&\text{PROPERTY}(P, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \\ &\text{PROPERTY}(Q, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \\ &\text{PROPERTY}(R, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM})\end{aligned} \tag{41}$$

This last move we can think of as analogous to partial execution in prolog or to some sort of compilation. Here, P, Q and R are the different properties which may be associated by various relations (such as "flattened") with the WORM for "lemon". It now seems a straightforward step to combine these unconditional constraints into a single unconditional constraint as follows.

$$\text{PROPERTY}(P, x) \vee \text{PROPERTY}(Q, x) \vee \text{PROPERTY}(R, x) \Rightarrow \text{DESCRIBE}(x, \text{LEM}) \tag{42}$$

It seems that the constraint we now have is of pretty much the same sort as the one expressed by the attribute-value structure in (2)[12]. What is important to note, though, is the important intuitions that have been lost in trying to reconstruct prototype theory from the Relational View. Above all else, it seems, this is what makes the Relational View a more attractive proposition than either classical theory or prototype theory. So what are these important intuitions?

Firstly, the moral of coherence requires that something like theories should underly word use. This is what COWORMs are intended to express. Having abstracted over them, as prototype theory seems to, we lose the possibility of respecting the important arguments made by Murphy & Medin, among others. Secondly, we have had our family of conditional constraints replaced by a single unconditional constraint. This violates the unitary assumption which appeared to lead to an unfortunate circularity: we argued earlier that this assumption would lead us to represent one meaning inside another. It seems that this is what has happened in obtaining a representation like (2). The fact of the matter is that we have lost the generalisations that a word like "flattened" expressed in the Relational View. In prototype theory, this information concerning flattening is subordinated to the level of a feature in a data structure, such as the one for "lemon", for example. Assuming such structures describe word meanings, it seems, then, that we have been led to represent meanings inside meanings. Thirdly, we now require some additional apparatus to respect the distinction between central and peripheral exemplars. This is precisely what the weights on the disjunctions in (2) achieve. Such weights, though, are not required under the Relational View, and we might require additional arguments to provide their justification. And fourthly, there is

[11] Strictly, we do not have to assume that B_1-B_3 hold but simply that so many background conditions can be realised or made explicit that the remaining background conditions always hold. Alternatively, we may conclude that there are no background conditions to be satisfied.

[12] Ignoring the fact that (42) has the power to express what have been called feature co-occurrence restrictions which (2), as it stands, does not.

the moral of Context Dependence. In the formalisation of the Relational View, the contribution of context was made relatively explicit. The constraints which licence the application of a word to an object were all conditional. That is, they depended on various background conditions holding. Whereas we were not, nor should we attempt to be, explicit about the nature of these background conditions, we were explicit about their role in word meaning. Our reconstruction of prototype theory has left the role of context at best inexplicit, at worst ignored. Such a result can only be unsatisfactory.

I have tried to argue above that though we *can* reconstruct prototype theory from the Relational View, this could only be done through doing great violence to our intuitions. Another argument that is of significance concerns Wittgenstein's (1953) much quoted observations on "games".

7.2 On Family Resemblance

Wittgenstein's characterisation of the category of games as having a family resemblance structure has often been taken by prototype theorists as motivation for their view. This is perhaps not surprising when we consider that prototype theory is an example of what we may call "cluster-concept" views. For the notion of "family resemblance" has been taken by many, Kripke (1972, p. 258), for example, to support such views. In this section we will suggest that, on the contrary, an equally plausible, and arguably better, account of family resemblance is inherent in the Relational View.

According to the Relational View, underlying the various uses of a word (and the objects it can describe) is not a single data structure or a single constraint or meaning relation, but a (possibly indefinite) number of such constraints. Now, while these constraints differ, importantly, they are all related. In particular, they are all related to some particular core constraint. In the case of Putnam's lemon, for example, the constraint underlying the various uses of "lemon" (for normal and for flattened lemons) are, as we have seen, systematically related. This fact is expressed by the use of what we have called COWORMs. Crucial for our concerns is the fact that such relations may belie the appearance of non-monotonicity or defeasibility. The specification of oval shape, for example, may be defeated by the WORM for "flattened". And, at least in principle, as the Puzzles of section 2 have shown, any such specification may be so defeated. This fact, then, allows that the various constraints underlying the uses of a given word apply to objects which, though related, may not necessarily exhibit common properties. It is this same fact, then, which allows us to claim that these constraints form a family. Thus, under the Relational View, family resemblance among the members of a category is captured by the idea that the various uses of a word belie a *family of constraints*. That is, a family of different but related meanings.

Now, if one takes Wittgenstein's arguments to support the Unitary assumption, then his concern with family resemblance can only be taken to support some means of making more complex, or more flexible, the sort of data structure underlying the meaning of a word. That is, the Unitary assumption, together with the observation of family resemblance, seems to force the adoption of some "cluster-concept" position. Such is prototype theory. However, a rejection of the Unitary Assumption, leads to a denial of the connection between the observation of family resemblance and "cluster-concept" views. The former need not lead to the latter. Indeed, we have seen this with the Relational View.

It should be obvious that the intention is not, in this meagre fashion, to attempt any kind of exegesis of Wittgenstein's intentions. Rather, simply to point to the possibility of an alternative account of family resemblance. So, to recap, indeed there may be no common features underlying the use of "game", for "game" has no *single* meaning. Rather, there is a family of different but related meaning relations underlying the various uses of "game". That is, a *family of constraints*.

7.3 On Concepts

We saw in section 7.1 that prototype theory, ostensibly a theory of concepts, can be reconstructed by "tinkering" with the idea of a family of constraints or meaning relations which is central to the Relational View. What prototype theory calls word meanings, though, seem best described by single, unconditional constraints on word use. Accepting the Relational View suggests that prototype theory cannot be a theory of word meaning. The suggestion which is most important for the purposes of this section is that, even as a theory of concepts, prototype theory is unsatisfactory.

The morals that we drew in section 3 are drawn on the basis of psychological evidence concerning theories of *concepts*. As we have seen, prototype theory appears to transgress a number. The morals of coherence, representational economy and context dependence, in particular, are not respected. We have argued that the Relational View does not suffer a similar fate. But the Relational View we have developed might, at first glance, seem a strange candidate for a theory of concepts. The reason is as follows.

Our comparison of prototype theory and the Relational View relied on the fact that the role played by such a structure as (2) in prototype theory is played by attunement to a family of constraints in the Relational View. This may seem disconcerting since people have been happy to think of concepts as single entities just as in (2). Replacing this notion, we have suggested that the work of a single concept is done by the attunement to a family of constraints. Indeed, whether a word applies to a given object, is determined not by some fixed entity but by some process of combining fixed entities called WORMs. So, under the Relational View, there is no single (mentally represented) object that plays the role of what has traditionally been called a concept. Rather, the traditional notion of a concept would, in these terms, amount to an abstraction over a whole family of mentally represented objects each of which expresses attunement to individual constraints or meaning relations. Though the conclusion that there may be no single object corresponding to a concept may strike some as perplexing, it is, fortunately, quite consistent with some recent work on concepts by Barsalou (Barsalou, 1987).

8 Conclusion

In this paper I have attempted to develop a psychologically plausible account of word meaning. The psychological literature suggests certain morals that should be respected. I have tried to show how the two major approaches, classical and prototype theory, variously transgress these morals. By paying attention to the considerations of coherence, representational economy, central and peripheral exemplars and context dependence a framework can be developed to respect these morals. The Relational View is an attempt to do just this. While the core meaning of a word can be described by a data structure or WORM, the peripheral meanings can be seen to be given by combinations of these data structures or COWORMs. Since these structures are constructed in context, the picture we have painted is one that requires word meanings to be strongly context sensitive. This position seems to be supported by the discussion and very influential arguments of Clark (1983). Indeed, the construction of COWORMs in context seems to be a form of the *sense creation* he argues so strongly for. We have also seen how the simple relational framework is able to give a general account of the problems of word meaning that we considered earlier as well as the problems of the meaning of kinship terms that the recent analyses of Lakoff and Macken have raised. Finally, from the perspective of the Relational View, I have suggested a number of implications. One amounts to a reinterpretation of prototype theory. By starting from the Relational View and trying to reconstruct prototype theory from it, it can be seen that prototype theory does violence to a number of intuitions. Another concerns the notion of family resemblance which, I have suggested, is better accounted for by the notion of a *family of constraints* than by the notion of a "cluster-concept". Finally, as a theory of concepts, prototype theory seems unsatisfactory. Indeed, it seems that the Relational View is better placed to account for recent work on concepts that suggests their illusory and transient nature. Also, as we have seen, what the

Relational View considers to be descriptions of word meanings, are far removed from what have traditionally been thought of as concepts. That is, from the perspective of the relational view, it seems there is a clear distinction to be drawn between concepts and word meanings, a distinction that is not often drawn in the psychological literature. Importantly, this distinction has emerged quite naturally from arguments drawn, in essence, from the psychological literature itself.

References

Barsalou L. W. (1982). Context-Independent and Context-Dependent Information in Concepts. *Memory & Cognition, 10,* 82-93.

Barsalou L. W. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In Neisser U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 101-140. Cambridge: Cambridge University Press.

Barwise J. (1985). *The Situation in Logic-II: Conditionals and Conditional Information.* CSLI report no. CSLI-85-21.

Braisby N. R. (1990). Prototypes and Word Meaning. In Dunbar G. L., Franks B. W. & Myers T. F. (eds.), *Papers from the 1989 Edinburgh Round Table on the Mental Lexicon.* Edinburgh Working Papers in Cognitive Science, Volume 4. Centre for Cognitive Science, University of Edinburgh.

Clark, H. H. (1983). Making Sense of Nonce Sense. In d'Arcais G. B. F. & Jarvella R. J. (eds.). *The Process of Language Understanding.* Chichester: John Wiley and Sons. 297-331.

Cohen B. & Murphy G. L. (1984). Models of Concepts. *Cognitive Science, 8,* 27-58.

Cohen G. (1983). *The Psychology of Cognition.* London: Academic Press.

Cresswell M. J. (1978). Semantic Competence. In Guenthner F. & Guenthner-Reutter M. (eds.) *Meaning and Translation: Philosophical and Linguistic Approaches,* 9-27. London: Duckworth.

Fodor J. A., Garrett M. F., Walker E. C. T. & Parkes C. H. (1980). Against Definitions. *Cognition, 8,* 263-367.

Hull C. L. (1920). Quantitative Aspects of the Evolution of Concepts. *Psychological Monographs, 28.*

Kripke S. A. (1972). Naming and Necessity. In Davidson D. & G. Harman (eds.). *Semantics of Natural Languages.* Dordrecht: Reidel.

Lakoff G. (1987). Cognitive Models and Prototype Theory. In Neisser U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 63-100. Cambridge: Cambridge University Press.

Macken B. (1990). Out of the Mouths of Babes. This volume.

Medin D. L. & Wattenmaker W. D. (1987). Category Cohesiveness, Theories and Cognitive Archeology. In Neisser U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 25-62. Cambridge: Cambridge University Press.

Murphy G. L. & Medin D. L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review, 92,* 289-316.

Nunberg G. D. (1978). *The Pragmatics of Reference.* PhD Thesis. Reproduced by the Indiana University Linguistics Club.

Osherson D. N. & Smith E. E. (1981). On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition, 9,* 35-58.

Putnam H. (1975). Is Semantics Possible? Chapter 8 in *Mind, Language and Reality, Volume 2: Philosophical Papers.* Cambridge: Cambridge University Press.

Putnam H. (1978). Meaning, Reference and Stereotypes. In Guenthner F. & Guenthner-Reutter M. (eds.) *Meaning and Translation: Philosophical and Linguistic Approaches,* 61-81. London: Duckworth.

Quine W. V. O. (1960). *Word and Object.* Cambridge, Mass.: MIT.

Rosch E. H. (1973). On the Internal Structure of Perceptual and Semantic Categories. In Moore T. E. (ed.) *Cognitive Development and the Acquisition of Language.* New York: Academic Press.

Rosch E. H. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General, 104,* 192-233.

Rosch E. H. & Mervis C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology, 7,* 573-605.

Rosch E. H., Mervis C. B., Gray, W., Johnson D. & Boyes-Braem P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology, 8,* 382-439.

Seligman J. (1990). Perspectives in Situation Theory. This volume.

Smith E. E. & Osherson D. N. (1984). Conceptual Combination with Prototype Concepts. *Cognitive Science, 8,* 337-361.

Smith E. E., Osherson D. N., Rips L. J. & Keane M. (1988). Combining Prototypes: A Selective Modification Model. *Cognitive Science, 12,* 485-527.

Vygotsky L. S. (1962). *Thought and Language.* MIT.

Wittgenstein L. (1953). *Philosophical Investigations.* Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.

Prototypes and Word Meaning

Nick Braisby

1 Introduction

This paper arises from a concern over the bifurcation of the psychology and philosophy of language on the issue of natural kinds. Its objective is to formulate a position on word meaning which is both psychologically and philosophically plausible. In the next section I shall outline some conceptions of meaning that have figured in the psychological literature. In the following section I shall draw parallels with conceptions of meaning found in the philosophical literature. The striking nature of these parallels makes it all the more distressing that current philosophical arguments suggest that the psychological conceptions of meaning are not conceptions of meaning at all. In making this point, I will make much use of Kripke's arguments in which the analytic is equated with the necessary *a priori*. After outlining Kripke's position, I shall suggest that a way of preserving the intuitions behind the various psychological positions is to view the Kripke-Putnam argument as having the force of a *reductio* and to regard the analytic as something other than the necessary *a priori*. In particular, I will pursue the possibility that analytic truths may be *contingent*. I will suggest that there are two assumptions of the traditional approach which need to be questioned. One concerns the issue of whether meaning relations need be necessary or may be contingent, the other concerns the assumed unitary nature of meaning. This last concerns the assumption prevalent in both psychological and philosophical approaches that words have *single* meanings. Whereas suggestions such as these may appear to divorce psychological conceptions of meaning even further from philosophical ones, on the contrary, I believe it offers their potential wedding. Firstly, I shall suggest that the philosophical theory of situations developed by Barwise & Perry, gives us the alternative conception of analyticity that we require. Secondly, I will argue that a more radical version of what Clark calls sense creation allows us to construe words as having multitudinous meanings. Both of these notion are crucial in developing a framework for word meaning, an example of which I will briefly outline in section 8.

2 Some Psychological Conceptions of Word Meaning

In their article, *What some concepts might not be*, Armstrong, Gleitman & Gleitman briefly detail the most prominent conceptions of word meaning that have figured in the psychological literature[1]. One, perhaps not surprisingly the most prevalent, is the view that the meanings of common nouns can be described in terms of singly necessary and jointly sufficient conditions. Such a view seems successful for certain nouns, generally ones applying to concrete, artificial categories. For example, one might regard having a certain computational power as a necessary, though not a sufficient, condition for a pocket calculator. Hence, the possibility that "pocket calculator" means, among others, an entity with a certain computational power. This view has perhaps been exemplified by the work of Katz & Fodor (1963). Their semantic markers amount to definitions of the meanings of words in terms of necessary and sufficient conditions. A related view is that of Fodor et. al. (1980) who hold that the meanings of words may be described in terms of meaning postulates. "Bachelor", for example, may be related via meaning postulates to "unmarried man". Again, these can be seen as expressing necessary and sufficient conditions. However, the major problem for this conception of meaning is the apparent difficulty to isolate any necessary and sufficient conditions for certain nouns, particularly those that label natural kinds.

[1] That such conceptions have often been confused with views of concepts will not concern us. Here I am interested just to outline views that have been held, regardless of whether they are confused.

The problem of isolating necessary and sufficient conditions underlying the meaning of a word reduces to the problem of identifying common features or properties, these features or properties being jointly sufficient for the application of that word. The existence of common features is, however, a hypothesis and one that has suffered much criticism. Wittgenstein, in his *Philosophical Investigations* (1953), implores us to try to find common features which might underly the meaning of "game" and, from the failure of this enterprise, concludes that none exist (§66, §67). Instead, Wittgenstein claims that those entities to which a term applies can be likened to a family. Entities are members of the family not because of their possession of certain common properties but rather because each and every member of this family exhibits a high degree of family resemblance. That is, there is assumed to be a collection of properties, often termed a cluster, associated with the family, possession of a certain number of which entails family membership. This powerful idea has been taken to vindicate the notion that such clusters describe the meanings of words and recent expositions of prototype theory are among the most explicit formulations of this position. It is this conception which today stands out among psychological conceptions of word meaning.

There have been numerous attempts both to suggest that prototype theory can account for compositional meaning and that it cannot and many are not content with the notion of prototypes and cluster-concepts at all (cf. Armstrong, Gleitman & Gleitman, 1983). However, in a long series of articles, Smith & Osherson and others have advanced the conjecture that the meaning of adjective-noun combinations may be describable in terms of the composition of the constituent prototypes (cf. Smith & Osherson, 1981; Osherson & Smith, 1984; Cohen & Murphy, 1984). Their model is perhaps one of the most explicit attempts to formalise the notion of prototypes. Prototypes are represented in terms of attribute-value matrices where values of attributes may be disjoint and weighted. That is, the prototype for "lemon", for example, may specify an attribute of shape but the values this attribute may take may be disjoint between perhaps, flat, oval or whatever and values may be weighted indicating that some values are more likely and more indicative of lemonhood than others. Attributes also receive a weighting indicating their diagnosticity. This last complication need not detain us and is omitted from the example which follows.

$$\left[\begin{array}{l} \text{shape:}\left\{\begin{array}{l}\text{oval:9*}\\ \text{square:1}\end{array}\right\} \\ \\ \text{colour:}\left\{\begin{array}{l}\text{yellow:8*}\\ \text{green:2}\end{array}\right\} \\ \\ \text{taste:}\left\{\begin{array}{l}\text{acidic:8*}\\ \text{sweet:1}\end{array}\right\} \\ \\ \text{isa:fruit:9*} \\ \vdots \end{array}\right] \qquad (1)$$

The attribute-value structure above applies to a large number of entities to which the word "lemon" applies. The fact that values are disjoint suggests that there need be no features common to those entities and such a suggestion invites the conclusion that prototype theory has formalised Wittgenstein's notion of family resemblance. However, this is only apparently, and not strictly, true. In order that the prototype, represented above, can apply to an entity it is necessary that the entity in question is either oval or square in shape, either yellow or green in colour and so on. These conditions are deemed to be jointly sufficient since prototypes are taken to admit all and only the entities in the associated category[2]. What we have then is similar to what we had before: a word's meaning can be described in terms of certain necessary and sufficient conditions. What is different is that these conditions are now complex and disjoint which means we need see no (simple) common features. Whereas some may argue that this point is trivial and amounts to a misunderstanding of the spirit of the cluster-concept position, on the contrary, I will argue in section 4 that this is precisely why prototype theory must necessarily fail to be a theory of meaning.

[2]This is not to suggest that membership is all or none. Prototype theory admits of entities which may be poor instances of a category. But the issue of how *good* a category member a particular instance may be is separate from whether it *is* a member.

3 Some Philosophical Conceptions of Word Meaning

It is probably fair to say that the current domination of the psychological literature on word meaning by cluster-concept views, belies a division. Many theorists hold what may be called *hybrid* views: that the meaning of a word consists of two components, one being a core describable in terms of necessary and sufficient conditions and the other being some set of identifying conditions describable in terms of a cluster concept. The suggestion is that one way of interpreting prototype theory is as a theory of the determination of reference. The fact that there is general disagreement on the proper interpretation of prototype and related theories betrays current confusion in the field. One way of trying to make more sense of the two conceptions of meaning on offer is by exploring the recent philosophical literature on word meaning. Here, as we shall see, the dominant notion is that of analyticity.

Philosophers have puzzled over proper names for some considerable time and their musings are particularly fertile ground for anyone interested in reaping the benefits of semantic analysis. Russell's theory of proper names[3] has been influential (Russell, 1956). In opposition to Mill, who held the intuitively appealing view that proper names have denotation but no connotation, Russell held that proper names abbreviate definite descriptions. "Aristotle", we assume, might abbreviate the definite description, "the teacher of Alexander" and, in this world, such a description certainly appears to pick out Aristotle and noone else. Now this thesis is not unlike that advanced in the psychological literature before the ascendancy of prototype theory. Our theory of common nouns like "calculator" might hold that its designation is determined by the applicability of some description such as "has such-and-such computational power". However, just as with that theory of common nouns, Russell's theory of proper names also came under attack and it is perhaps not surprising that one line of attack came from those proposing the notion of a cluster-concept.

One problem with Russell's approach to which Strawson objected was that of explaining the common determination of reference: different people having different knowledge of and beliefs about Aristotle who may be unaware of the fact that Aristotle taught Alexander, nonetheless seem to refer to the very same man with utterances of "Aristotle". This apparent fact about language use suggests a problem with Russell's position, for if a proper name abbreviates a definite description then who is to say which description it abbreviates or what may happen when people take the name to abbreviate different descriptions? For Strawson, these problems are not due to the spirit of Russell's enterprise but to its finer details. The suggestion is that Russell's theory needs an added degree of sophistication, in particular, the notion of a cluster-concept.

Under Strawson's view (Strawson, 1959), associated with every proper name is a set or cluster of propositions, a "presupposition-set", whose truth is presupposed by the use of that proper name. His theory of proper names is then that the referent of a proper name is that unique individual which satisfies some "reasonable proportion" of the propositions in this set. This proportion need not be fixed, nor indeed need the membership of the set be given in any precise way (pp. 191-192). A similar position is adopted by Searle in his article *Proper Names* (Searle, 1958). Whereas for him proper names such as "Aristotle" don't have Fregean senses, Searle, like Strawson, holds that referring *uses* of proper names do nonetheless imply the satisfaction of a certain number of statements commonly attributed to the bearer of that name. Aristotle, whoever he may be, the argument goes, *necessarily* satisfies some certain proportion of the true statements commonly attributed to him[4]. Such cluster-concept views are not so different from the position advanced in prototype theory. The various disjunctions in (1) can be viewed as clusters of propositions each of which need not be satisfied by an individual lemon, but necessarily some proportion being satisfied.

The relations between these approaches to proper names and the psychological literature on natural kind terms are striking. It is the nature of these relations that make it both particularly unfortunate and of great relevance to psychological debate that the philosophical literature has generally come to reject these approaches to proper names. The rejection has as its base the arguments of Kripke and Putnam, focussing on *necessity*, a word used much in this discussion but to which scant attention has been paid.

[3] I use "proper name" in the conventional sense and not in the sense of Russell's logically proper names.

[4] Searle goes on to say "any individual not having at least some of these properties [...the ones commonly attributed to Aristotle...] could not be Aristotle"

4 Necessity and Word Meaning

Kripke's seminal article *Naming and Necessity* is an attempt to ground discussions of meaning firmly within the bounds of the analytic tradition. A number of intuitions concerning analyticity have prefigured Kripke's discussion. Leibniz distinguished between "truths of reason" and "truths of fact" and further suggested that the former have the property of being necessary while the latter are contingent. Leibniz also made the suggestion that necessary truths of reason are ones that are true in "all possible worlds". Leibniz' distinction parallels a distinction made by Hume between "relations of ideas" and "matters of fact" and it is this distinction which is oft thought to lie between analytic and synthetic truths. Kant's concern with this distinction lead him to outline a number of characteristaions of analytic truths. For him, all statements could be cast into sunject-predicate form and analytic statements were ones whose predicate was contained in and identical to their subject. Consequently, analytic truths are ones which cannot be denied without expressing self-contradiction. In recent times great attention has been focussed on the validity of the distinction. Quine (1951), White (1950) and illustrious others have sought to argue that the distinction is without foundation for natural languages, while others have argued in opposition. In this section, I am concerned not so much with this last debate but with Kripke's characterisation of analytic truths as ones which are both necessary and *a priori*. In particular, it is the explication of meaning in terms of necessity which is of central concern.

Kripke seems to a conception of analyticity not unlike Leibniz'. For Kripke, analytic statements are those statements which are true (or false) in all possible worlds and, hence, necessarily true (or false). Since analytic statements depend for their truth or falsity on "relations of ideas" or, less prosaicly, on their meaning they are also deemed to be *a priori*. Thus, the analytic is the necessary *a priori*. This criterion on statements of meaning, however, is too strong for Russell's theory of proper names to withstand. Returning to Aristotle, of whom it is only contingently true that he taught Alexander, we can see that the statement "Aristotle is he who taught Alexander" is not analytic since it does not express a necessary truth: we can envisage possible worlds in which Aristotle never even met Alexander. Consequently, "Aristotle" cannot mean "the teacher of Alexander". This then is the house that Kripke built and Russell's theory cannot be admitted but neither, it seems, can the cluster-concept views.

In pitting Kripke's arguments against the cluster-concept views of meaning, we must remind ourselves that these views can be seen as attempts to express necessary and sufficient conditions on meaning. Searle, for example, claims it is *necessary* that "Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him". But from Kripke's point of view this is not the case. Just as possession of each of those properties can only be contingent so is possession of any proportion of them. Consideration of counterfactual conditionals seems enough to convince us of this. "If Aristotle had not taught Alexander then such-and-such" does not suggest that Aristotle would no longer be Aristotle. Yet this is precisely what follows under the Searle-Strawson views if we strengthen the antecedent to incorporate all the properties commonly attributed to Aristotle. Kripke's point is that Aristotle is still Aristotle (i.e., the same man) even under possible circumstances where he has none of the properties we normally associate with him. He is the same by *fiat*, by stipulation: by the very act of using the name "Aristotle" we refer to Aristotle. In Kripke's terms, "Aristotle", as with other proper names, *rigidly designates*.

Kripke's arguments strongly suggest that the standard description and cluster-concept accounts of proper names are neither theories of meaning nor of the determination of reference. The scope of Kripke's argument is, however, not nearly so narrow and his attention turns to natural kind terms where his arguments have perhaps greatest import for psychological conceptions of word meaning.

If (1) describes the meaning of "lemon" then certain statements, which exemplify the contents of (1), should be analytic. In particular, "Lemons are fruit and either yellow or green and either oval or flat and either acidic or sweet" should express an analytic truth. According to Kripke's view, the truth of the statement should therefore be *a priori* knowable and it should be *necessary*: in Kripke's terms it should be true in all possible worlds. Evidently this statement is not analytic and the fact that we can envisage lemons of all hues, shapes, sizes, tastes shows this. Each property represented in (1) holds only contingently and so does their disjunction. This though is not necessarily a problem for a *theory* of prototypes. It can always be argued that the *actual* prototype for "lemon" is not as represented by (1). The question is whether prototypes can *in principle* give rise to statements that are analytic and truths that are necessary.

One way we can arrive at necessary truths from disjoining contingent ones is by expressing the law of excluded middle. For our case of lemons, we could have descriptions such as "yellow or not yellow". Now it is clear that this description applies to all (possible) lemons and so is necessary. The point about such descriptions, however, is that while capturing necessity they fail to capture sufficiency which is the other side of the coin of word meaning: "lemon" must not only apply to all lemons but it must apply only to lemons and not to non-lemons. The point being that tautologous descriptions such as "yellow or not yellow" are grounds for nothing: they in no way discriminate between lemons and non-lemons. An alternative route to necessary truths via the disjoining of contingent ones is by exhausting possibilities. For "lemon", for example, we could exhaustively include all the possible colours lemons might have. A suitable description might be "yellow or blue or green or red or black or...". Again, however, this description is of little utility in a theory where we require the meaning of "lemon" to distinguish between lemons and non-lemons. Any coloured object will satisfy this description.

We have considered two possible moves prototype theorists might make to counter the Kripkean arguments both of which capture necessity only at the expense of sufficiency. The "meanings" that would be left us would be ones which apply to objects indiscriminately. Such moves seem, therefore, out of the question. The problem we face now is the one which provides the motivation for this paper: philosophical and psychological conceptions of meaning are divergent. The divergence seems, at first glance, irreconcilable. In the next section I shall outline an argument which I believe offers a potential reconciliation, an argument that I believe has probably already been countenanced by Putnam, from whose article the section heading is shamefacedly stolen.

5 Is Semantics Possible?

Putnam (1975) raises the issue of the very possibility of doing semantics because of the apparent failure of a thesis which lies at the heart of the analytic tradition. The thesis is such as to presume that if a word has a meaning then there must be some statement which is analytic in virtue of that meaning. That is, the truth or falsity of the statement depends only on the meanings of its components and their mode of combination rather than on any empirical investigation. It is this thesis that drives the endeavours of Kripke and Putnam. That words such as "lemon" and "tiger" have meaning does not seem controversial. That there are statements which are analytic in virtue of those meanings does.

The first part of *Is Semantics Possible?* is occupied by two concerns. One is to suggest that the then current theories of meaning were in reality nothing of the kind; the other is to pursue analytic truths. The first concern is argued in much the same way as Kripke does in his criticisms of description theories of proper names. Necessity is the key and it is the failure of those earlier theories to express *necessary* truths concerning the meaning of natural kinds that renders them theories of something other than meaning. Putnam's suggestion, it seems, is that they are theories of which core facts it may be whose articulation allows the meaning of a word to be commonly grasped. Such a conclusion seems to offer little hope of reconciling psychological and philosophical views and it is the success of Putnam's own approach to meaning which must come under scrutiny. So it is the second concern of *Is Semantics Possible?* that is of more direct interest.

Putnam's search for necessary, analytic truths concerning the meaning of natural kinds leads him to consider a number of putative definitions of the meaning of his example, "lemon". Putnam's arguments serve to convince that each such definition unfortunately fails the test of necessity and so neither expresses any analytic truths nor captures the meaning of "lemon". It is the failure of this enterprise that raises for Putnam the spectre of the impossibility of doing semantics. It is the fact that standard assumptions of the analytic tradition are called into question by the failure to find analytic truths that prompts his extremely disturbing question.

The failure to find necessary truths concerning the meaning of natural kind terms suggests a number of responses. The move adopted by Kripke and Putnam is to suppose that this failure is to do with our *knowledge* of natural kinds. The fact that we have not discovered necessary truths is then a fact of our epistemological condition and not evidence for the non-existence of such truths. In this regard the argument

is not unlike that of Grice & Strawson (1956). In response to Quine's pessimism over the existence of the analytic-synthetic distinction (Quine, 1951), they claim that Quine's arguments concern our ability to make the distinction well-defined and not the matter of whether the distinction exists. Similarly, Kripke and Putnam assume that such necessary truths do exist, it is merely that they remain, in large part, undiscovered. While this essentialist response is open, it is one which allows little role for the psychology of language users to play in theories of meaning (cf. Franks, 1989). It is for this reason that it appears an unsatisfactory option for a theory which hopes to draw psychology back into the philosophical fold.

The Kripke-Putnam move to essentialism is, however, not the only possible response to the apparent failure to find necessary truths. Another is to view Putnam's dilemma as being due to the fact that his conclusion has the force of a *reductio* on the premisses of his argument. It is to this alternative interpretation of the Kripke-Putnam position that I now wish to turn.

The fact that necessary truths are difficult to find and may, in certain cases such as natural kinds, be beyond our grasp can be taken to be such an unsatisfactory conclusion that the arguments on which it is based must be incorrect. In the case of the Kripke-Putnam arguments, I want to suggest not that the structure of the argument is wrong, for I think it is right, but that one premiss is at least in need of reconsideration. The premiss that analytic truths are also necessary ones is one, I believe, that cleaves psychological conceptions of word meaning from philosophical ones. Accordingly, I want to suggest that analytic truths need not be necessary ones. In the next section I will outline the analysis of meaning offered by Situation Theory and argue that meaning so construed provides the basis for this different conception of analyticity. In the following section, another assumption, concerning the assumed unitary nature of meaning, is discussed. Rejection of this assumption paves the way for a view of word meaning which is briefly presented in section 8.

6 Meaning in Situation Theory

Kripke's arguments rely heavily on the equation of analyticity with the necessary *a priori*. My argument in this section is not with the characterisation of analyticity as *a priori*. Instead, I shall be concerned to show that statements true in virtue of their meaning need not be necessarily true and that it is this characterisation of meaning which has lead to the bifurcation between philosophical and psychological approaches.

In the Relation Theory of Meaning of Barwise & Perry (1983), meaning is explicated in terms of constraints which, when actual, relate situations. The abstract notion of a constraint exists as a relation between situation types. Though Barwise & Perry don't offer a complete taxonomy of constraints, they certainly offer enough for our purposes.

Barwise & Perry consider two broad classes of constraint: unconditional and conditional. Within each category we can further distinguish between nomic, necessary and conventional constraints. A nomic constraint might be expressed by the statement that "smoke means fire" and is similar to Kripke's category of the contingent *a posteriori*. Necessary constraints might be expressed by statements such as "kissing means touching" and is similar to Kripke's category of the necessary *a priori*; that is, for Kripke, such a statement would express an analytic truth. Conventional constraints can be expressed by statements like "*rain* means rain". The fact that such and such word has been linked by language to such and such part of reality is naturally merely a fact of convention and also corresponds to Kripke's contingent *a posteriori*.

Cutting across this classification is the distinction between unconditional and conditional constraints. Barwise (1985) argues strongly for the value of conditional constraints in accommodating various puzzles and in order to illuminate the distinction between unconditional and conditional constraints we will use one of his examples.

Barwise tells us that when his daughter, Claire, rubs her eyes this means she is sleepy. Attunement to this constraint is what allows Jon and Mary-Ellen to get Claire to bed at appropriate times. The theory allows us the choice of whether we model this constraint as unconditional or conditional. If it is an unconditional constraint then it can be described as follows: *whenever* Claire rubs her eyes she is sleepy. The constraint that the Barwise's are attuned to is hardly likely to be of this nature, however. Were an abundance of pollen

to be in the atmosphere, even a fully alert Claire might start rubbing her eyes. It is this contingency that would lead Kripke to complain that rubbing one's eyes cannot *mean* that one is sleepy[5]. However, Barwise & Perry offer us an alternative. We can model the constraint as unconditional. That is, rubbing one's eyes means that one is sleepy *provided* that prevailing background conditions define a certain type of situation.

Before elaborating on the benefits of this way of looking at meaning, it is perhaps illuminating to relate it to some rather older proposals. Fillmore (1982) discussed the problems of defining the meaning of terms like "bachelor". Though certain individuals (Popes, gay men, muslims with less than the allowed number of wives) are technically bachelors to whom the term does not easily apply, there is no requirement to question the nature of the definition of "bachelor". The reason is that the meanings of words can be taken to be defined with respect to some set of background conditions. The fact that a word may not easily fit individuals to which the definition may apply may not be a fact about the meaning of the word but rather a fact about the background conditions with respect to which the word can be said to have its meaning.

6.1 Analyticity in Situation Theory

In characterising analyticity, we are naturally concerned with linguistic meaning and not the sorts of natural meaning discussed so far. However, situation theory still allows the choice of whether we analyse linguistic meaning in terms of conditional or unconditional constraints. Unconditional constraints appear to require necessity as a property of meaning. Now it seems from the preceding discussion that necessity is too strong a criterion. It prevents psychological evidence from having a bearing on philosophical notions of meaning. It also appears to tie semantics too close to ontology. Quine, for example, seems to adopt an ontology which excludes necessary truths altogether (Quine, 1953). So the positing of necessary truths really is an issue of one's ontology, not of one's semantics. One way of keeping semantics further from ontology is to take seriously the notion that meaning can be explicated in terms of contingent truths. This is precisely what situation theory's notion of conditional constraint offers. Accordingly, in section 8, we will pursue the notion that analytic statements are ones whose truth, in spite of being *a priori* knowable, is nonetheless contingent.

7 The Unitary Nature of Meaning

Having accepted the idea that meaning may be related to contingent truths, it is important to see that this does not of itself relieve us of our problem with natural kind terms. The prototype represented in (1) clearly embodies contingent truths. However, it has only limited applicability to the category of lemons. As previously argued, in order to apply to all lemons the prototype is likely to be rendered indiscriminate. What is required is a way of retaining the specificity of meaning while capturing the flexibility of meaning over different contexts.

Related to this is work which largely accepts the description theory of proper names. The work centres on the problem of trans-world identification. Roughly, on description theories of proper names, the bearer of that name may not, in certain possible worlds, satisfy the description. Under such circumstances, conditions must be given by which the name still refers to that individual. One response to capture the applicability of a name across worlds would be to divine necessary truths underlying the name. Another might be to treat the name as abbreviating a disjunction of contingent truths in the spirit of the cluster-concept approach. The generality of applicability of names across worlds renders the last suggestion untenable as the name becomes indiscriminate just as natural kind terms do in prototype theory while the assumption that there are necessary truths is perhaps one that should be justified on ontological rather than semantic grounds.

An alternative way of capturing flexibility in meaning while retaining specificity (i.e., while retaining discriminating meanings) is via what is variously known as Sense Selection or Sense Generation. The distinction is due to Clark (1983). According to Clark, novel uses of words can involve cases where they can be used to mean something other than their conventional meaning. As theorists we have a choice as to the status we accord these new meanings. We may suppose that the word is ambiguous having multiple meanings or

[5] Of course, we are not talking of linguistic meaning here.

we may suppose that meanings are, in some sense, created at the time of novel use. The latter is what characterises Sense Creation.

Returning to Kripke, we might want to argue that what Kripke is observing is more akin to a case of polysemy rather than the ineffability of meaning. In the case of "lemon", for example, we might argue that a conventional sense need only apply to some specified subset of lemons and that in the case of other lemons, related senses are generated. This proposal seems to advance a more radical version of Sense Creation, what we call Sense Generation which is expanded upon in more detail elsewhere (Myers, Franks & Braisby, 1989). In the next section I outline an alternative view of the meaning of natural kinds which utilises these two cornerstones of a combined philosophical and psychological approach: conditional constraints and sense generation.

8 A Relational View of Word Meaning

In this section I will briefly outline a view of word meaning which attempts to respect not only the considerations raised in this paper but also a number of considerations from the psychological literature that we have not discussed. A more detailed picture can be found in Braisby (1989).

The considerations of section 6 allow us a choice in describing the meaning of, say, "lemon". Both of the prominent psychological conceptions of meaning can now be thought of as precisely that; conceptions of *meaning*. It is this fact which allows a host of psychological evidence to bear on the choice and some of this evidence is discussed in more detail in Braisby (1989). However, what is clear is that the move taken by prototype theory, a move ostensibly to avoid contingency in meaning, is unnecessary. At least, it is unnecessary from the point of view of meaning. Considerations like this lead me to describe the meaning of "lemon" in what we might call pseudo-classical terms. I hope the following example will clarify this.

$$\begin{bmatrix} \text{shape:oval} \\ \text{colour:yellow} \\ \text{taste:acidic} \\ \text{isa:fruit} \end{bmatrix} \tag{2}$$

Let us suppose that this is related to the meaning of "lemon" and for convenience refer to all such Word Meanings as WORMs. WORM s then relate words (in this case "lemon") to properties (acidic taste, etc.). To say that this describes *the* meaning of "lemon" is, however, misleading because, as we shall see, the assumption of sense generation allows us to view "lemon" as having multitudinous meanings. Before turning to that consideration let us try and put the proposal we have on a more formal footing. In particular, we will model these WORMs in situation-theoretic terms as conditional constraints. The previous structure (2) then can be seen as describing the following constraint.

$$S_1 \Rightarrow S_2 \mid B \tag{3}$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P, \dot{p}; 1 \gg] \tag{4}$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{LEMON}; 1 \gg] \tag{5}$$

and

$$B = [\dot{b} \mid \dot{b} \models \ll \text{WORM}, \text{LEMON}, P; 1 \gg] \tag{6}$$

and

$$P = [\dot{p}_1 \mid \ll \text{SHAPE}, \dot{p}_1, \text{OVAL}; 1 \gg \wedge \ll \text{TASTE}, \dot{p}_1, \text{ACIDIC}; 1 \gg \wedge \ll \text{COLOUR}, \dot{p}_1, \text{YELLOW}; 1 \gg \wedge \ll \text{ISA}, \dot{p}_1, \text{FRUIT}; 1 \gg] \quad (7)$$

Without attempting a detailed explanation of this formalism[6] we can see roughly what is meant by (3). It tells us that a relation holds between types of situation S_1 and types of situation S_2, *provided* that certain background conditions prevail. The nature of the relation is such as to say that if we have a situation of type S_1 then we have a situation of type S_2 given, of course, that the background conditions prevail. These background conditions can be taken to define another type of situation B. (4) and (7) define S_1 as the type of situation in which (2) can apply to an object. In this case, the type of situation where an object has the properties of being oval in shape, yellow in colour, acidic in taste and being a fruit. (5) defines S_2 as the type of situation in which this same object can be described as a lemon, that is, in which the word "lemon" may apply to the object. What (6) tells us that this relation (3) only holds when the properties the object has are related to the word "lemon" by the WORM relation.

The fact that we model word meanings in terms of conditional constraints allows us to reject the Kripke-Putnam arguments. However, we still have a problem, one which is directly analogous to that faced by those who hold to description theories of proper names. The fact that descriptions are contingent means that, in certain cases, they will not satisfied by the very bearers of the proper name they abbreviate. If "Aristotle", for example, abbreviates "the teacher of Alexander", how can the proper name refer to Aristotle in (counterfactual) circumstances where Aristotle never taught Alexander? There are numerous moves one can make to avoid this problem. One, due to David Lewis, is to enrich one's ontology to include counterparts (Lewis, 1968). Such a move accepts that reference to an individual in another possible world is not reference to that individual in our world but rather to one of that individual's *counterparts*. Accordingly the fact under description theories proper names seem not to refer to the same individual across worlds is not a problem. It only makes sense to talk of the *same* individual in the same world. What should be clear is that our proposal concerning natural kinds now faces an exactly analogous problem. How do we refer to lemons in circumstances where the properties represented in (2) do not hold? The answer to this problem, as we suggested earlier, comes from an approach derivative of Clark's Sense Creation called Sense Generation.

To illustrate sense generation, consider a flattened lemon. Such an individual will neither satisfy (2) nor the antecedent condition in (3). Consequently, if (3) was our only option, we would be unable to describe such an individual as a lemon. The proposal though is that we can generate other senses for "lemon", derivative on the WORM for "lemon", which in turn describe different but related constraints to (3) which will allow such maverick individuals to be described as lemons.

The proposal is that one way of generating an appropriate sense is to *combine* knowledge of "flattening" with our knowledge of "lemon". "Flattening", for example, seems to express a relation between things of indeterminate shape and things of flat shape. By allowing (2) to stand as one of the relata of this relation we can determine the other relata which, intuitively, describes flattened lemons. Combinations such as these we will label COWORMs. COWORMs thus generate further senses for given words that are related to the standard sense describable in terms of WORMs. These COWORMs also express constraints and the example that follows describes the COWORM which arises from the combination of "lemon" and "flattened".

$$S_1 \Rightarrow S_2 \mid B' \quad (8)$$

where

$$S_1 = [\dot{s} \mid \dot{s} \models \ll P_2, \dot{p}; 1 \gg] \quad (9)$$

and

$$S_2 = [\dot{s} \mid \dot{s} \models \ll \text{DESCRIBE}, \dot{p}, \text{LEMON}; 1 \gg] \quad (10)$$

[6] This is something that is best left to experts and Barwise & Perry (1983), Barwise (1985) and Barwise (1989) are where the interested reader will find help.

and

$$B' = [\dot{b} \mid \dot{b} \models \ll \text{WORM, LEMON, } P_1; 1 \gg \wedge \ll \text{WORM, FLATTENED, } P_1, P_2; 1 \gg] \quad (11)$$

and

$$P_2 = [\dot{p} \mid \ll \text{SHAPE, } \dot{p}, \text{ FLAT; } 1 \gg \wedge \ll \text{TASTE, } \dot{p}, \text{ ACIDIC; } 1 \gg \wedge \ll \text{COLOUR, } \dot{p}, \text{ YELLOW; } 1 \gg \wedge \ll \text{ISA, } \dot{p}, \text{ FRUIT; } 1 \gg] \quad (12)$$

Roughly, this can be read as follows. If an individual has the properties of flat shape, acidic taste, yellow colour and is a fruit then it can be described as a lemon *provided* those properties are given by the combination of the WORMs for "lemon" and "flattened".

9 Conclusion

I have argued that psychological conceptions of word meaning are strikingly similar to philosophical ones and that this fact makes it all the more distressing that recent philosophical arguments leave no role for psychological evidence to play in theories of meaning. In particular, the arguments of Kripke and Putnam which reduce meaning relations to those that hold between words and some essential properties of the individuals to which they apply cleave psychological and philosophical conceptions of meaning wide apart. One reason for this, I have suggested, is that Kripke and Putnam may be making too strong an ontological commitment and explicating semantic issues in these terms. The adoption of necessity as underlying analyticity is a case in point. The existence of necessary truths is an ontological committment which is open to doubt. An alternative way of analysing meaning is to allow analytic statements to be contingent. Such a conception is one allowed by the recent theory of situations developed by Barwise & Perry. This allows us to rescue what we might call description theories of natural kind terms. However, it leaves us with an analogous problem to that of trans-world identification which faces those who hold description theories of proper names. Adopting an ontological solution to this problem is one which seem unsatisfactory. Indeed, Kripke's argument can be seen as one which replaces a strong and controversial ontological committment to counterparts with a weaker, more acceptable one to necessary truths. Our criticism of Kripke prevents us from adopting an ontological solution. This problem is one to which a suggestion made by Clark to do with the creation of senses is readily amenable. By adopting a more radical position, sense generation, we can allow words to mean differently according to circumstances. The process of sense generation employed relies on the combination of what Clark might call pre-stored senses and what we have taken to underly WORMs. COWORMs, as we have called them, are the result of sense generation. The fact that such COWORMs apply to individuals to which the corresponding WORM may not offers the hope that we have a semantic solution to the traditional problem facing description theories. As described, the mechanism of combination can clearly not be unconstrained. The formulation here allows that a variety of contextual factors may play a constraining role. Clearly, the identification of these factors and others which may constrain the process of sense generation is a matter for considerable future research. What the research described in this paper has attempted to do is simply demonstrate that such research may be of philopsophical as well as psychological significance.

References

Armstrong S. L. L. Gleitman & H. Gleitman (1983). What some Concepts might not be. *Cognition, 13,* 263-308.

Barwise J. (1985). *The Situation in Logic-II: Conditionals and Conditional Information.* CSLI Report No. CSLI-85-21.

Barwise, J. (1989). *The Situation in Logic* . CSLI Lecture Notes 17. University of Chicago Press: Chicago, Illinois.

Barwise J. & J. Perry (1983). *Situations and Attitudes.* Bradford Books, MIT Press.

Braisby N. R. (1989). Situating Word Meaning. Paper presented at the Conference on Situation Theory and its Applications, Asilomar, California. March, 1989.

Clark, Herbert H. (1983). Making Sense of Nonce Sense. In d'Arcais, G. B. F. and Jarvella, R. J. (eds.). *The Process of Language Understanding.* Chichester: John Wiley and Sons. 297-331.

Cohen B. & G. L. Murphy (1984). Models of Concepts. *Cognitive Science, 8,* 27-58.

Fillmore C. (1982). Towards a Descriptive Framework for Spatial Deixis. In Jarvella R. J. & W. Klein (eds.). *Speech, Place and Action: Studies in Deixis and Related Topics* . Chichester: John Wiley and Sons. 31-59.

Fodor J. A. Garrett M. F. Walker E. C. T. & C. H. Parkes (1980). Against definitions. *Cognition, 8,* 263-367.

Franks B. W. (1989) Criteria and Concepts: an Anti-Realist approach to Word Meaning. PhD. Thesis. Centre for Cognitive Science, University of Edinburgh.

Grice H. P. & P. F. Strawson (1956). In Defense of a Dogma. *The Philosophical Review, 65.*

Katz J. J. & P. Postal (1964). *An Integrated Theory of Linguistic Descriptions* . Cambridge, Mass., MIT Press.

Kripke S. A. (1972). Naming and Necessity. In Davidson D. & G. Harman (eds.). *Semantics of Natural Languages.* Dordrecht: Reidel.

Lewis D. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy, 65,* 113-126.

Myers, T. F. B. W. Franks & N. Braisby. (1989) Partiality and Coherence in Concept Combination. Forthcoming in the Proceedings of the International Colloquium on Cognitive Science, San Sebastian, Spain. May 1989.

Osherson D. N. & Smith E. E. (1981). On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition, 9,* 35-58.

Putnam H. (1975). Is Semantics Possible? Chapter 8 in *Mind, Language and Reality, Volume 2: Philosophical Papers* . Cambridge: Cambridge University Press.

Quine W. V. O. (1948). On What There Is. *Review of Metaphysics* .

Quine W. V. O. (1951). Two Dogmas of Empiricism. *The Philosophical Review, 60* .

Russell B. (1956). On Denoting. In March R. C. (ed.). *Logic and Knowledge* . London: Allen and Unwin. 41-56.

Searle J. R. (1958). Proper Names. *Mind, 67,* 166-73.

Smith E. E. & Osherson D. N. (1984). Conceptual Combination with Prototype Concepts. *Cognitive Science, 8,* 337-361.

Strawson P. F. (1959). *Individuals* Methuen, London.

Wittgenstein L. (1953). *Philosophical Investigations.* Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.

White M. (1950) The Analytic and the Synthetic: An Untenable Dualism. In *John Dewey: Philosopher of Science and Freedom* . New York, The Dial press.

Sense Generation
or
How to Make the Mental Lexicon Flexible

Bradley Franks Nick Braisby

1 Introduction

In this paper, our aim is to consider some possibilities for certain aspects of a theory of word meaning, and thereby to motivate what we call *Sense Generation*. We briefly outline the crucial phenomena of flexibility and specificity of senses in Section 2. In Section 3 we sketch some of the sorts of object which might be required in a theory of sense, and, most importantly, the relations between them. This provides the apparatus for a discussion of some possible theories of senses, in Section 4. In Section 5, we turn to a more detailed exposition of Sense Generation, indicating how it may account for flexibility and specificity. Finally, we sketch some implications of this view and touch on some wider concerns for theories of word meaning.

2 Preliminaries

The phenomena of flexibility and specificity are best illustrated by example.

> Mary is giving a dinner party at her home in the country. Unfortunately, her cupboards are bare. The appetites of her voracious guests are, however, whetted by the sight of Mary's pet mouse, Midge, tucking into some mouse food, and by the sight of Mary's pet canary eating some bird food. Aware of all this food-eating, one of Mary's less subtle guests asks "Do you have any food, Mary?". Mary replies that there is none but proceeds to feed her dog, Mungo.

How are we to resolve the semantical nature of the guests' problems? They are led to believe that Mary has no food and yet they can clearly see that she has: that she has pet food, but no food fit for human consumption. We require that any theory of word meaning respect the intuition that there are different but related senses attached to the word *food*, senses, for example, which apply to different types of food: human food, dog food, bird food, etc. We employ "sense" in a similar manner to Clark (1983): as the mentally represented aspects of the semantic content of a word on a particular occasion of use; we will be more precise about this in 3. The fact that the same word can seemingly have many different senses, illustrates what we call *diversity*. One aspect of the diversity of senses is illustrated by the fact that *food* seems to have senses corresponding to both types and subtypes of food. In this case, it has senses ordinarily associated with *pet food*, and its subtypes, for example *mouse food*. This aspect of diversity is what we call *specificity*: some senses of a word appear to be more specific than others. The fact that *food* may have senses for different types, for example, senses for "mouse food", "dog food", "mouse food", etc., illustrates another aspect of diversity, *flexibility*.

Our discussion of theories of sense will concentrate on several factors: we will be concerned to respect the arguments of, among others, Clark (1983) and Murphy & Medin (1985), which we will outline in more detail later. We also require that any plausible account of reflect two important distinctions between types of senses: between senses that express default information and those that express non-default information, and between senses of vague and ambiguous words. Throughout, our overriding concern is to provide an evaluation which not only respects basic linguistic intuitions, but does so in a way that is consonant with a broad range of psychological considerations.

3 Some Components of a Theory of Sense

In this section, we will set out some aspects of a theory of sense. Two categories of object play a central role in our discussion. The first, *Senses*, are descriptions that we take to mediate relations between uses of words and their referents. These descriptions are both publicly specifiable and mentally representable. The notion of sense as employed here, although derived from that of Frege, does not carry a commitment to Frege's abstract semantical "third realm", distinct from the realms of mental and physical objects. The most important aspect of senses for our purposes is the way in which they guide linguistic behaviour. The application of a word to an entity (objects, events, substances - any individuum) is mediated by the sense of that word: in particular, the description that constitutes the sense subsumes the description of the object. So the uses of a word must be explicable in terms of the sense or senses which that word possesses. In this way, senses may be taken to classify the linguistic behaviour of agents.

The description of the phenomena of flexibility and specificity relied upon the various senses noted for *food* being different. This assumption was motivated by the application of what Evans (1976) labels the "intuitive criterion of identity" for senses. This determines that if a rational agent can both assent to and remain agnostic about the application of a referring expression to an entity when used in utterances of the same sentence, then that referring expression must have two different senses. As an illustration, reconsider Mary's dinner guests. Here, *food* is being used in different ways: sometimes it is being used to refer to all food, and at other times to types of food. So it is possible that one of Mary's guests could both assent to, and dissent from the statement, "Mary has food in her house". So it is quite felicitous for Mary to say, of the same entity (i.e., some mouse food, say), both that it is food, and that it is not. The intuitive criterion of identity for senses then requires that we treat *food* as having such different senses. Different senses express the fact that an entity may have different "modes of presentation" with respect to an agent: different ways the agent may refer to that entity. They also correspond to different ways of cognizing that entity: they are indicative of different perspectives that an agent may adopt.

The second type of object that we require is *lexons*. Most accounts of the psychology of language presuppose the existence of a "mental lexicon", in which words have "entries", that contain orthographic, phonological, morphological, syntactic and semantic information. The semantic component has been variously referred to as a "concept" or "lexical concept"; in order to avoid correlative unwarranted assumptions, we will refer to it as the lexon. A lexon, then, is a description that defines the stable mental representation in this mental lexicon; it also forms the semantic contribution of a word to the meaning of the expressions of which it forms a part. We also assume that senses are derivative in some way on lexons. That is, language users arrive at a sense for a word through first accessing its lexon. Given the multiplicity of senses which a word may have it is clear that a major problem for theories of is the relation between senses and lexons. Our discussion of such theories rests primarily on the way this issue is addressed.

Pre-theoretically, we are led to believe that senses usually outnumber the words with which they are associated: that is, the senses of a given word always number one or more. Considering the relations between words and their senses in terms of lexons then gives rise to two crucial questions. One concerns the *number* of lexons we postulate in order to effect these relations; and the other concerns the relations between the *contents* of senses and lexons.

In order to facilitate our discussion we will describe the contents of lexons and senses in terms of feature-structures like the following, which may describe the lexon for *chair*:

$$\begin{bmatrix} \text{legs:} & \begin{bmatrix} \text{number:} & 4 \end{bmatrix} \\ \text{seat:} & \begin{bmatrix} \text{number:} & 1 \end{bmatrix} \\ \text{made-of:} & \text{wood} \end{bmatrix} \tag{1}$$

This feature-structure is not intended to be an exhaustive specification of the content of the lexon for *chair*; it is presented for illustrative purposes only. If any lexon or sense has the same feature-structure as this one, we may conclude that they are in fact identical lexons/senses. There may be cases where one feature-structure subsumes another, by having the same content or some addition of features. Another possible case is where

two feature-structures cannot be ordered by this relation. The former is indicated by the relationship between the structures for *chair* and *arm-chair*, and the latter by that between *chair* and *rocking chair*:

$$\begin{bmatrix} \text{legs:} & \begin{bmatrix} \text{number: } 4 \end{bmatrix} \\ \text{seat} & \begin{bmatrix} \text{number: } 1 \end{bmatrix} \\ \text{arms:} & \begin{bmatrix} \text{number: } 2 \end{bmatrix} \\ \text{made-of:} & \text{wood} \end{bmatrix} \qquad \begin{bmatrix} \text{legs:} & \begin{bmatrix} \text{number: } 0 \end{bmatrix} \\ \text{seat} & \begin{bmatrix} \text{number: } 1 \end{bmatrix} \\ \text{arms:} & \begin{bmatrix} \text{number: } 2 \end{bmatrix} \\ \text{rocker:} & \begin{bmatrix} \text{number: } 2 \end{bmatrix} \\ \text{made-of:} & \text{wood} \end{bmatrix} \tag{2}$$

The subsumption relation amounts to the kind of relation that holds between a type and one of its tokens; typically, we might assume that that defining features of a type are possessed by a token that can be categorised as a member of that type.

On just these two dimensions, the number and contents of lexons and senses, we distinguish three classes of theory. The first, Strong Sense Selection ($\mathcal{S}$), is that which Clark demonstrated to be unsound; we will note some additional problems. $\mathcal{S}$ may appear to be a straw man; a more plausible alternative is the second class, Weak Sense Selection ($\mathcal{W}$). $\mathcal{W}$ has two variants, both of which appear to be flawed. The third class, Sense Generation, avoids these difficulties, and is the one we would like to endorse.

4 Problematic Theories of Sense

4.1 Strong Sense Selection

Two assumptions identify $\mathcal{S}$. Firstly, the number of lexons: there is a lexon for each and every sense of a given word. Secondly, the contents of the lexons and senses: $\mathcal{S}$ assumes that the contents of each lexon and its corresponding sense are identical. $\mathcal{S}$ accords well with standard model-theoretic analyses of word meaning. For example, the approach taken in Montague-style semantics requires that different interpretations for the same syntactically unambiguous linguistic string result from the same word having different basic expressions. In the case of *bank*, for example, there would be two distinct basic expressions, $bank'_1$ and $bank'_2$, in the lexicon. $\mathcal{S}$ then offers the possibility of being able to treat the diversity of senses which might be associated with *food* in the same way, and thus to provide a precise semantics. All of the idiosyncratic information which demarcates senses is thus represented in lexons.

The diversity of senses that might be attached to *mother* provides another illustration of $\mathcal{S}$. Such senses include "adoptive mother", "biological mother", "surrogate mother", "foster mother" and "step-mother". According to $\mathcal{S}$ each of these senses is assigned a distinct lexon whose content expresses that of the sense.

Despite its prevalence in formal approaches to word meaning, $\mathcal{S}$ has some irremediable deficits. Some of these, relating to flexibility, coherence and ambiguity, are also problems of $\mathcal{W}$ and we will turn to these in 4.3. However, there are also problems unique to $\mathcal{S}$. One concerns the number of lexons we are led to hypothesise in order to capture diversity. This is essentially the same point that Clark made in respect of "nonce" senses; however, as we have seen, even for common nouns such as *food* and *mother* the number of senses greatly exceeds the number of words. Accordingly, in $\mathcal{S}$, so does the number of lexons. This is problematic since the multiplication of lexons must make psychological sense. It is unclear that this is so in the case of $\mathcal{S}$ since such a multiplication places an intolerable burden on memory and presumably would result in a highly complex search procedure. These problems are difficult enough in the case of the interpretation of single words: in the case of combinations, such as simple noun phrases, there would be an explosion of combinatorial possibilities, in which the appropriate sense would have to be selected from a list comprising each and every permutation of all of the lexons associated with each constituent.

Regarding flexibility, $\mathcal{S}$ appears to proffer a solution that treats vagueness and context-sensitivity in the same

way as ambiguity: by postulating independent lexons for each sense of a vague word. We will return to this in 4.3. It is also clear that specificity raises difficulties for $\mathcal{S}$. The issue is whether we can have a limitless number of ever more specific senses for a given word. We remain agnostic about this possibility though it is clear that very many senses may be associated with the same word. However, $\mathcal{S}$ rules out the possible unboundedness of specificity by *fiat*. That is, the only way that $\mathcal{S}$ can possibly capture specificity is via the multiplication of lexons, and, given the uncontroversial assumption of a finite lexicon, the possible unboundedness of specificity could not be captured by $\mathcal{S}$.

$\mathcal{S}$ also appears unable to distinguish between senses which express default information and those which express non-default information. For instance, the default sense for *mother* is, presumably, "biological mother" yet this sense is accorded the same status as the other senses of *mother*. That is, they are each assigned a separate lexon. Of course, $\mathcal{S}$ theorists may have in mind some other bit of theoretical apparatus which to capture this distinction. The fact is, though, that as it stands, $\mathcal{S}$ does not respect this very important distinction. An alternative to this rather straw-mannish way of trying to capture the phenomena is offered by Weak Sense Selection.

4.2 Weak Sense Selection

$\mathcal{W}$ is characterised by three assumptions: firstly, there may be more than one lexon for a given word; secondly, there may be more than one sense for a particular lexon. Thirdly, it is assumed that the contents of senses and the contents of corresponding lexons lie in the relation of subsumption: that is, the only possible difference between a sense and the lexon from which it comes, is that the sense may have had features added. $\mathcal{W}$ is more appealing than $\mathcal{S}$ in the following ways. Firstly, senses seem to be intrinsically context sensitive: in conventional circumstances *mother* has the sense of "biological mother", but in a social work inquiry, for example, *mother* may have the sense of "biological mother who is also a carer". A way in which this context-sensitive specificity can be captured is through some process by which features are added to the contents of lexons in a manner appropriate to context. Though $\mathcal{W}$ does not specify such a process it is clearly implicit in its definition. This aspect of $\mathcal{W}$ is in the spirit of the findings of Barsalou (1982): different senses may be different "context-dependent" elaborations of a single "context-independent" lexon.

Another appeal of $\mathcal{W}$ is the fact that it allows lexons to express generalisations with respect to the category to which a word applies. That is, $\mathcal{W}$ allows that *lion* may have various senses but that the lexon for *lion* may be a description that applies to all (and only) lions. This is again quite appealing given standard assumptions about word meaning. The arguments of Kripke (1972) and Putnam (1975) for example, assume that senses apply to all and only those individuals to which the word applies. Further, the fact that $\mathcal{W}$ distinguishes between the different senses deriving from a lexon, that is, between those whose content is the same as the lexon and those whose content is an elaboration of the lexon's, *may* allow for the expression of default information. That is, default information might be expressed as part of the content of lexons.

There are two extreme versions of the $\mathcal{W}$ thesis: one is that the number of senses and lexons are equal, which forces equivalence with $\mathcal{S}$; another is what we might term the "Generality" option. This results from the assumption that a word has only one lexon, whose content may be added to and made more specific. $\mathcal{W}$ allows that the number of lexons may be intermediate between the number of words and the number of senses: what we will term an "Intermediate" option. Our previous discussion of $\mathcal{S}$ allows us to consider just the Generality and the Intermediate options of $\mathcal{W}$. The Generality option would operate in the following way. *Mother*, for example, would be assigned a single lexon whose content would subsume all the senses that *mother* can have. So all of its senses result from the addition of features to this lexon's content. Given the diversity of senses for *mother* such a lexon must needs be maximally unspecific. In contrast, the Intermediate option allows *mother* to have more than one lexon underlying its senses. For example, we might have lexons for "biological mother" and "surrogate mother", say. The latter might be further specified to yield senses for "adoptive mother", "step-mother" and so on.

There are a number of problems with $\mathcal{W}$ some of which we will deal with in 4.3. However, we will outline some problems unique to $\mathcal{W}$ here. A critical problem of the Generality option is that it appears to be unable to express default information. Reconsider the example of *mother*. The appropriate lexon cannot express a relation of genetic inheritance because although some mothers are related this way to their children and some are not (e.g., foster mothers). To specify such a feature in the lexon would be to exclude mothers such as these

from the domain of application of the lexon and all its associated senses, since features can only be added to the lexon and not taken away. For similar reasons, the lexon cannot express any relation of caring between mothers and children. And so on for any other feature-specification which we might ascribe to the lexon for *mother*. Arguments such as these indicate that in many cases the Generality option leaves us with a maximally unspecific lexon. This, however, flies in the face of the strong intuition that words *do* have default senses. The fact that this option renders the expression of such defaults as difficult to obtain as the expression of exceptions is a major deficit.

The Intermediate option postulates a *certain* multiplicity of lexons: there might be more than one lexon for *mother* (in contrast to the Generality option), but less than would be postulated by $\mathcal{S}$. The critical problem here is exactly how the number of lexons might be determined. Whereas for $\mathcal{S}$ and the Generality option, there is an overt constraint on the number of lexons postulated, it is not clear what principle there could be for deciding on the number of lexons in Intermediate option. For example, what lexons might we postulate for *chair* so that it may have a sense corresponding to "rocking-chair"? Since feature-adding is the only way in which senses may be derived from lexons, the sense for "rocking-chair" shown in (2) can only be derived from a lexon which either does not specify the number of legs or specifies no legs. Under this Intermediate option we are allowed to postulate several lexons for *chair*: one corresponding to (1), say, and one corresponding to that for "rocking-chair" in (2). A problem may arise, however, in the case of a special type of rocking-chair having no legs and no rockers. Assuming that *chair* can have this sense, the question arises as to how it is derived from the lexons we have postulated. If we only have feature-adding at our disposal, such a sense simply cannot be derived from either of the lexons for chair we have postulated. Our only option is to suppose that there is another lexon for chair. The problem is that, in principle, there seems to be no bound to the number of exceptional chairs we can imagine and for each type, we would be led to posit an additional lexon. The issue then would be, what degree of exception do we rule out as invoking a new lexon? One way of constraining the number of lexons might be to determine a threshold for permitted specificity of lexons: if a particular sense is more specific than the threshold level, it must be represented as a sense deriving from a particular lexon (and not as a lexon in itself). Three problems render such a criterion untenable.

Firstly, it is not clear just how we could go about comparing the relative specificity of senses that have non-overlapping contents; for example, is "biological mother" less specific than "adoptive mother"? The operation of this criterion is perspicuous within groups of senses that *can* be ordered according to specificity (i.e., where the only difference between senses is in the degree of specification of the same set of features), but not in groups that cannot be so ordered. The general application of such a criterion would require a complete theory of content for lexons and senses, and some precise and motivated means for comparison. Neither are at present available. Secondly, to stipulate that lexons must be relatively non-specific may mean that they cannot express default information, since this is typically quite detailed and specific in nature. A third problem concerns the plausibility of postulating independent lexons to account for senses that are discriminable though related. This will be picked up in the next section.

4.3 Difficulties with Sense Selection

There are three principal flaws common to Sense Selection accounts. The first concerns the multiplication of lexons. The second concerns the ability of Sense Selection to account for the full range of flexibility. And the third concerns the underlying commitment to monotonicity, which gives rise to the first two problems.

The difficulty with assuming multiple lexons is that it is not clear to what extent they plausibly reflect mental representations. Arguments from linguistics and psychology caution against unprincipled multiplication. The linguistic considerations concern the difference between ambiguity and vagueness. Accounting for the diversity of senses by postulating distinct underlying lexons assumes that they are, synchronically, wholly independent. That is, supposing different lexons for a given word assumes they are as different as different lexemes with a single orthographic/phonological form. Postulating two lexons for *food* ("animal food" and "human food") treats *food* as an ambiguous item like *bank*. This amounts to making no distinction between different senses of a vague term, and different senses of an ambiguous term. However, if we consider any of the standard linguistic tests for ambiguity (Cruse, 1986), then we find that the independence of content assumed by postulating different lexons does not hold for examples such as *food*. For example, consider zeugmatic contexts (those which give rise to two different senses of a word at one time): "He sat on the bank whilst fishing and put his

cheque in it". The strong contrast or opposition between the two senses of *bank* requires the postulation of independent lexons to account for them. In contrast, the various senses for *food* or *mother*, noted earlier, do not produce an opposition of sufficiently marked character. The examples are, rather, characterised by the relatedness of the various senses: they are distinct but clearly *not* independent.

Multiplication of lexons is also countered by psychological considerations raised by Murphy & Medin (1985). Murphy & Medin's discussion bears on the issue of the mental representations underlying the application of words to referents. In the current framework, these are lexons. The question for both $\mathcal{W}$ and $\mathcal{S}$ is whether the postulation of multiple lexons accords with psychological evidence. Murphy & Medin's arguments convince us that such multiplication is unwarranted. Consider whether the postulation of independent lexons for "animal food", and "human food", say, is justified on psychological grounds. The thrust of Murphy & Medin's arguments is to suggest that categories such as food are highly structured, and that the application of *food* to individuals thereof is highly dependent upon this structure. That is, the application of a word to entities in a category reveals what Murphy & Medin call "coherence". And the crucial point regarding coherence is that the application of a word to such an entity is dependent upon our theories concerning that entity. It is the fact that entities can be related by theories that allows their grouping together to be psychologically plausible. If lexons are to be psychologically plausible, then entities that form a coherent category (e.g., all different types of mother) should all fall under the extension of the same lexon.

However, even if we were to allow some multiplication of lexons - that is, even if the above considerations have no purchase - there are still cases in which the postulation of independent lexons to underly senses for a particular word would be implausible. These include the "contextual expressions", discussed by Clark (1983), and many examples noted by Nunberg (1977). Nunberg notes several different possible referents (and therefore, senses) for *newspaper*: a particular token of the newspaper (as in, "here's your newspaper, sir!"), the newspaper company as a whole (as in, "the newspaper's profits are less than expected!"), and a particular journalist (as in the case of a dubious piece of governmental behaviour: "don't say a word, the newspaper is here!"). It is clear that we would not want to claim that *newspaper*, for example, has a pre-stored lexon that expresses the sense of the third use ("journalist"). Yet this is precisely what Sense Selection accounts would have to hypothesise, since this sense of *newspaper* could not be said to be a simple specification of a lexon for *newspaper* (as $\mathcal{W}$ might aver). Clark argues that the parsing of contextual expressions (including certain denominal verbs, such as *to teapot*) stems from the *creation* of interpretations associated with those phrases. It is clear that the same kinds of considerations apply equally to Nunberg's examples. Since the only possibility of a creative process for Sense Selection is provided by the specification mechanism of $\mathcal{W}$, the only contextual expressions that $\mathcal{W}$ could accommodate would be those that are mere specifications of pre-stored lexons. And this cannot capture the flexibility evidenced in, for example, denominal verbs. It is clear that, even though Clark's discussion of $\mathcal{S}$ might appear to have attacked a straw man, his arguments have a broader significance, and have played a central role in undermining the more plausible $\mathcal{W}$.

The preceding discussion leads to the conclusion that none of the versions of Sense Selection are adequate to the task of accounting for the phenomena in a way that does justice to basic psychological and linguistic intuitions and requirements. The major problems stem from Sense Selection's adherence to monotonicity. That is, to the assumption that any alteration in the content of a lexon in the formation of a sense must be feature-addition. The Sense Generation approach (section 5) circumvents these problems by denying precisely this assumption and then tracing the ramifications.

5 Sense Generation

In opposition to the above types of theory, Sense Generation regards the variation we observe in senses to be due not to variation in lexons nor the generality of lexons but to some generative process which generates various senses from a (lexon) base.

5.1 Aspects of Sense Generation

Sense Generation is characterised by the following assumptions. Firstly, the number of lexons is identical to the number of non-ambiguous words. A single linguistic string is assigned more than one lexon if and only if it has

genuinely unrelated senses, as indicated by tests for ambiguity. Secondly, the content of a lexon comprises the default sense of the word. Thirdly, the different senses of a word are generated from the lexon for that word. Fourthly, generation may result in a sense that is non-monotonically related to the lexon; that is, generation may result in a sense that does not simply add features to those of the lexon: features may be retracted or negated in the generation process.

Such a view can readily account for the kinds of example that are so problematic for Sense Selection. Recall the different senses for *newspaper*. In Sense Generation the lexon corresponds to the default sense of *newspaper*, perhaps as in "the newspaper hit the mat". The two other senses we have identified would then be non-monotonically derived from this lexon. The sense, for example, in "the newspaper's profits halved" would have to involve a retraction of those features expressing the facts that newspapers are material objects, made of paper, containing ink, etc. It would also need to include features expressing facts about businesses, finance, employment, etc. What Sense Generation claims is that this latter sense can indeed be generated from the lexon for *newspaper*. That is, there is some process by which features are negotiated. The precise nature of such processes is, clearly, a matter for further empirical enquiry but we suggest two possibilities. One is that argued for in Franks (1989) which involves the emergence of the features of a sense being constrained by some implicitly attached noun derived from an instantiation. For example, the sense of *fake gun* may be partly derived from the lexon for *replica* (thus adding features such as the degree of resemblance to a gun, and the way the object might be constructed), which is accessed as a result of the lexon for *fake* defeating certain features of that of *gun* (like firing bullets). Another is that suggested in Braisby (1989) where a related process of combining lexons results in the defeating of features. For example, the sense of *lion* which applies to stone lions, can be seen to result from the combination of a relational lexon such as "statue" and the lexon "lion". Similarly, for other non-default senses: they are derived from the combination of default and other lexons.

5.2 Implications of Sense Generation

The link between Sense Generation and Clark's *sense creation* should be noted. Clark's argument is that there is a restricted and well-defined group of contextual expressions, for which a sense creation process is necessary. Clark suggests that, for other kinds of expressions, a selection mechanism may be adequate. Since Clark's focus of attention is not conventional senses, the impression may be gained that these are not similarly contextual in nature. Where Sense Generation differs from sense creation is in its firm committment to the view that conventional uses are higly contextual and therefore require some generative process to explain the diversity of their senses. It follows that there is no clear dichotomy between contextual expressions and those used conventionally, rather there is a gradation.

Sense Generation is, in general, agnostic about the precise time-course of the role of *context* in determining senses. Indeed, it is meant as a formal, abstract characterisation of a class of theory. Consequently, it is compatible with more detailed accounts which suppose context to play a pre-access as opposed to a post-access role. That is, although context may choose from a number of possibilities for generation, it may also rule out certain possibilities prior to any generative process. Ultimately, this entails that exceptional senses need not require a longer time-course than default ones. Clearly, the exact time-course underlying the generation of senses is a matter for empirical enquiry: it suffices to note that Sense Generation is compatible with either outcome.

One implication of Sense Generation is that there are meanings of which senses are descriptions. Two questions arise: is there a meaning relation which the default sense of, say, *mother* describes? And, since there are many senses of *mother*, are there also many meanings to *mother*? Whereas the traditional theory of meaning may answer these questions in the negative, support for affirmative answers comes from *Situation Theory*. Indeed the notion of meaning relations as conditional constraints allows us to claim that default senses are descriptions of meanings. The assumption that there are many meaning relations underlying the uses of a word is also perfectly compatible with the framework of Situation Theory.

A further range of issues concerns the connection between Sense Generation and various hypothesised structures and contents for lexons. The Sense Generation approach allows us to endorse certain aspects of both *classical* and *prototype* representations, whilst rejecting problematic implications of both views. In terms of the epistemological rationale of the classical approach, Sense Generation rejects the search for common features or necessary conditions underlying the sense of a word (as a result of nonmonotonicity), whilst it allows us to

retain the economical representations that would result were the search for necessary conditions successful. In contrast, the relations between the various generated senses for a word mirror the intuitive and epistemological underpinnings of family resemblance (as a result of nonmonotonicity and the emergence of new features), which is acheived without postulating prototype representations, with their attendant difficulties. That is, the various senses generated for a word in different contexts will be related by a family resemblance, in line with Wittgenstein's (1953) original formulation.

Senses are descriptions that mediate a word's reference. There may appear to be a tension between this fact and the unlimited scope for nonmonotonicity in Sense Generation: if generation defeats all of the default features of a word's lexon, then we may refer to an entity through a description having nothing in common with the usual properties of the type of referent. This point is countered by considering the *perspectival-relativity of categorisation*. A situated agent, in referring to an entity through a particular sense, can be seen to be adopting a perspective on that entity. Even though Sense Generation allows of the logical possibility that words may refer to any manner of entity, an important constraint is deemed to operate. Namely, the agent in making such a reference must be adopting a particular perspective. While a theory of perspectives is something we lack, we note that for such a reference to be posited there must be independent evidence concerning the perspective adopted. Further, we suppose that the nature of the perspectives that people may adopt is such as to determine the content of senses which they relate to referents. This may of itself limit the degree of permissible difference between the content of a word's lexon and its senses, since one of the purposes of perspectives is to allow mutual reference. That is, one constraint is that several agents must be able to share a single perspective.

6 Conclusion

In this paper, we have sketched some of the assumptions that underly prevalent views of the senses of words (Sense Selection views). We have also considered some of the problems that arise when such views attempt to account for some basic phenomena of word meaning in a way that is consonant with linguistic and psychological desiderata. This then motivated an alternative view that circumvents these difficulties (Sense Generation). Finally, we noted some of the other advantageous implications of Sense Generation.

References

Braisby N. R. (1989). Situating Word Meaning. Paper presented at the Conference on Situation Theory and its Applications, Asilomar, California. March, 1989.

Clark, H. H. (1983). Making Sense of Nonce Sense. In d'Arcais, G. B. F. and Jarvella, R. J. (eds.). *The Process of Language Understanding.* Chichester: John Wiley and Sons. 297-331.

Evans G. (1976). *The Varieties of Reference.* Oxford: Basil Blackwell.

Franks B. W. (1989) Criteria and Concepts: an Anti-Realist approach to Word Meaning. PhD. Thesis. Centre for Cognitive Science, University of Edinburgh.

Murphy G. L. & Medin D. L. (1985). The Role of Theories in Conceptual Coherence. *Psych. Rev., 92,* 289-316.

Partiality and Coherence in Concept Combination

Terry Myers Bradley Franks Nick Braisby

1 Introduction

This paper addresses the problems of lexical representation and coherence. Our discussion of these problems leads to what we call a sense generation approach. We will present two accounts from within this approach and illustrate them with respect to concept combination. In section 2 of this paper we describe the tension that exists between the two phenomena of semantic flexibility and specificity. There are accounts of lexical representation in which this tension raises major problems and there are others in which it provides a clue to resolving these problems. The former are characterised by what Clark (1983) calls the selection of senses and the latter by what he calls sense creation. In section 3 we discuss the implications of sense selection accounts for coherence and concept combination, indicating problems that arise. We rely heavily on Murphy & Medin's (1985) arguments and suggest some extensions. In section 4 we present the sense generation view (which is related to Clark's notion of sense creation) which avoids such difficulties. Central to this account is a consideration of partiality, reduction of partiality being brought about by constraints provided by the discourse and situational factors. In section 5 we will sketch two different approaches to concept combination within this framework. In section 6 we draw conclusions and raise questions regarding the nature of coherence and lexical representation.

2 Flexibility and Specificity

Consider the following example:

> Rudy is at the ballet watching a single ballerina dancing on stage. Excitedly, he whispers to his friends, "Isn't she delightful?". His friend Ron agrees, replying, "Yes, she's very beautiful". His other friend, Reg, agrees too, saying, "Yes, what an exquisite pirouette". But Rudy replies, "I simply meant she's a wonderful person".

Delightful in the above exchange seems to mean different things for the various individuals involved. From Ron's perspective, *delightful* means 'physically delightful'; from Reg's, *delightful* means 'dancing delightfully', while for Rudy *delightful* means 'delightful as a person'. From the perspective of an usher walking past who only hears Rudy's first comment, the meaning of *delightful* is indeterminate with respect to these specific meanings.

The order of authorship is arbitrary. The first two authors were supported by European Basic Research Action, ESPRIT project 3175. The second two authors were supported by ESRC studentships Nos. COO428622024 and C00428722001 respectively. We would like to express our deep gratitude to Robin Cooper, George Dunbar and other members of the Mental Lexicon Workshop for many useful discussions and insights. We would also like to thank Richard Cooper for his help in preparing the printed version of this paper.

Suppose that we want to characterise the semantic component of the lexical entry for *delightful*. We might propose that the lexical entry is sufficiently general to encompass all the observed meanings. In which case, we require a lexical entry which is general enough to capture all the different meanings that can be conveyed by *delightful*, including those illustrated in the exchange. Even if we could specify such a general lexical entry for *delightful*, however, this would not then allow the possibility for *delightful* to convey, say, the meaning 'delightful as a person'. Yet, in use, *delightful* can convey just this. What is more, any of the meanings indicated above could be made even more specific if more contextually provided information were available. For example, Ron's meaning for *delightful*, 'physically delightful' could have been more specific: physically delightful with respect to coiffure' if we know that Ron is, say, a hairdresser. In principle, there is no bound to this degree of specificity and so a general lexical entry seems even more implausible. Tied to this idea of unboundedness, we also want to allow that meanings can be indeterminate: for example, the meaning for the usher, is non-specific with respect to the meanings for Rudy and his friends. In conclusion then, our treatment of the lexical entry as general, being motivated by the observation of flexibility, renders the lexical entry unable to provide for the specificity of meaning that we observe. Seemingly, we can't treat the phenomena of flexibility and specificity independently of one another: to do so produces the tension we have illustrated[1].

3 Sense selection

3.1 Weak and strong sense selection

For the purposes of a distinction to be made later, we will refer to the semantic contribution of the lexical entry as the "lexical concept", and the intended meaning as the "generated sense". One approach to characterising lexical entries is to treat the different meanings of a word as different lexical concepts. For example, under this view we might suppose that the lexical entry for *mother* includes several lexical concepts for *mother*, which we might label "foster mother", "step mother", and "biological mother" - in the same way as, say, *bank* is assumed to have a number of independent lexical concepts underlying its use. A strong version of this view would hold that for each and every different sense a different lexical concept is accessed from the lexical entry. A weaker version maintains that there may be independent, unrelated lexical concepts from which, on different occasions of use, different senses may be elaborated. For example, under the weak view, we could, in the context of a fairy tale, arrive at a sense for *mother* corresponding to "step-mother who is wicked" by elaborating the lexical concept for *mother* labelled "step-mother". Presumably, on the strong sense selection view we would select the appropriate sense, "step-mother who is wicked", relevant to the context and that could not be further elaborated.

Both versions of sense selection miss the crucial intuition that these senses are closely related; they are not at all like the different senses of *bank*. That is, *river bank* and *money bank* have clearly distinct meanings, and thus *bank* should have (at least) two different lexical concepts. In the case of *mother* the different lexical concepts we could hypothesise are clearly related. Accordingly, in the strong version, we have a granular picture of flexibility and specificity. Flexibility is granular in that the possible senses a word may

[1] A similar point has been made by, among others, George Dunbar. In Dunbar (1988) he suggests that, "while lexical meaning is flexible, lexical use can be very precise ... this is paradoxical"(p. 69).

have cannot alter and we have an impoverished view of flexibility. Specificity is also fixed by the set of senses in that senses can be no less or no more specific than those represented. The weak version allows senses to vary in terms of their specificity but we only have apparent flexibility: the range of possible senses is restricted by the initial choice of lexical concept.

3.2 Coherence in sense selection

In exploring the implications for coherence of the sense selection view we will treat Lakoff's (1987) analysis of *mother* within this account, although we accept that he may not actually be committed to such a view. That is, we might see each of his sub-models as a different lexical concept within the lexical entry for *mother*. A number of consequences then follow. In particular, we will argue that we cannot maintain both weak sense selection and a theory-based view of coherence. To see this we must turn to Murphy & Medin.

In Murphy & Medin's (1985) account we may distinguish between internal and external components of conceptual coherence. Both aspects derive from theories of the world. Internal coherence results from the relations between features for which the theory provides an explanation. For example, the various properties of chairs represented in the concept (its size, having a seat, its rigidity and so on) cohere in virtue of theories concerning the function of a chair. External relations between different concepts are also motivated by theories that support causal and explanatory connections between those concepts. Consider the concepts "cat" and "dog". According to Murphy & Medin these concepts are coherent to the extent that they are embedded in our theories of the world: the greater the number of connections between such concepts and the rest of our knowledge, the more coherent and stable they are. The connection of "cat" to many different aspects of our knowledge (such as our theories of domestic life, ownership, pleasure and so on) provides it with coherence over and above that which is provided by our theories of "cat" alone, which is the source of the concept's internal coherence. The same thing is true of complex concepts, for example "cat and dog". The concept is coherent as a result of there being many common theories of cats and dogs. Seemingly, on Murphy & Medin's view, we could not allow that complex concepts such as "elephants and lemons" or "stone lions and trout" have the same level of coherence as "cat and dog". This stems from the fact that such complex concepts has fewer common theories to relate the component concepts.

This view of the theories underlying coherence seems to be intimately connected to metaphysical issues. Accordingly, for Murphy and Medin, concepts result from the way that theories cut up the world. Consequently, categories that cut across ontological boundaries would not be very coherent. For example, "elephants and hopes" would not be considered coherent, unless this category were motivated by a theory. If this theory were consistent with one's other theories, the coherence of the concept would be enhanced. The claim of Murphy & Medin is that the flexibility of conceptual coherence derives from the flexibility of particular theories. On our interpretation of their claim, this means that there are two mechanisms for allowing coherence to be flexible. Either, in certain contexts, coherence is enhanced by constructing an explanation or theory or, alternatively, the theories that underpin coherence are inherently flexible. It is not clear how these mechanisms for flexibility in coherence would work for examples like the following.

> Imagine you are sitting on a park bench in London. In front of you is a stone statue of a lion. Consider how you would reply, if someone were to ask you, "Is that a lion?" Seemingly, there is one sense in which you might say "yes", and a related one in which you might say "no": and, given these senses, either reply could be appropriate, according to circumstances.

It seems, then, that the word *lion* can be used quite appropriately to talk of a statue of a lion. However, there is a question as to just how many of our theories of lions involve statues. Seemingly, our theories of lions are capable of cutting the world up so that we can categorically assert that a statue of a lion is not a lion. One approach to this phenomenon is to view it as a peculiarity of word use. However, we wish to preserve the intuition that there is an essential perspectival aspect to this case. That is, we can adopt a different perspective on a statue of a lion: we may simply view it as a lion. Additionally, in this case each of these perspectives seems equally coherent. As we see it, there are two possibilities for Murphy & Medin to account for this apparent flexibility in coherence. One is to invoke their suggestion regarding the flexible nature of theories. Another is to adopt the standpoint of weak sense selection, and thus hold that there are two lexical concepts for *lion*. In what follows we will explore the former possibility and conclude that the latter is their only plausible option. This in turn leads to an inconsistency.

Under the first option, that of invoking flexibility in theories, we envisage three possible mechanisms open to Murphy & Medin for achieving the flexibility of coherence. One mechanism that Murphy & Medin offer is that of the construction of an explanation in cases such as "elephants and hopes" where, seemingly, ontological boundaries are crossed. In the above case, apparently there is a similar crossing of ontological boundaries. Our theories of lions tell us that stone lions are not really lions, though presumably there are some theories of lions that are also theories of stone lions. Perhaps then it is that the coherence of the concept of lion in this case relies on a constructed explanation. However, since our stable, underlying theories concerning lions cannot all apply in the case of the stone lion, it must be that the use of lion in talking of a stone lion is less coherent than the use of lion in talking of a real lion. This follows because coherence in Murphy & Medin's view falls out of the number of explanatory links which structure the concept. Hence, on this view such constructed explanations necessarily lead to less coherent concepts than those stable explanations provided by the knowledge base. This is at odds with our intuitions. Seemingly both uses are, in this case, equally coherent. What we want is to allow that both ad hoc and stable explanations can provide for equal coherence. This is what we will argue for in section 4.2.

Murphy & Medin provide another mechanism to account for the flexibility in coherence that we observe. This is the possibility that we may have flexible theories. In the case of the stone lion, we might suppose that our theories concerning lions are flexible, accounting for the observed flexibility of coherence in use. The use of *lion* to talk of a stone lion can be coherent as we would want in virtue of the fact that the same theories concerning lions can also be theories concerning stone lions. However, this does not accord with the intimate connection that theories have to metaphysical concerns. If our theories concerning lions are also theories concerning stone lions then we have no principled manner of cutting the world such as to differentiate between non-lions (like our stone lion) and real lions. We take it that this undermines the metaphysical position that Murphy & Medin adopt.

A third possibility that we believe Murphy & Medin may allow for is the option of theory change. This option entails that in order to preserve the coherence of the lion concept when employed to talk of a stone lion, the theories underlying the concept would have to change. This, however, seems problematic. Since they adopt the Quinian position of viewing theories as components of a large and intricate web, a change in one will have ramifications for the others. In this case we would have to change our theories concerning lions so as to allow the possibility of inanimate lions. Changes as radical as this, as they acknowledge, would require a global reorganisation of one's knowledge base. It would also require, under most interpretations, a divorcing of theories from metaphysics. This is precisely what we argue for. In section 5 we will present two views that are not tied to the assumptions that lead to these counter-intuitive consequences. As we have seen there are good reasons to suppose the kind of flexibility observed in coherence is not to be captured by the notion of flexibility in theories. We now turn to the second possibility we pointed to earlier.

The alternative position that we believe Murphy & Medin might adopt is the one that we call weak sense selection. That is, the above arguments will not go through if we assume that there are two lexical entries and hence two lexical concepts for *lion*. Aside from the arguments already presented against this option, it seems there are further considerations that would prohibit Murphy & Medin from taking this view. Following Quine (1960), we might like to distinguish between two lexical entries for *light*. This move seems legitimate. After all, it seems that the theories we have for light-weight objects do not have to be the same as those for light-coloured objects. For the case of *lion*, although lions and stone lions are differentiated by the theories provided by the knowledge base, lions and stone lions are nonetheless very closely related. It is presumably the case then that since we can relate lions and stone lions we have theories to do just this.

Returning to our interpretation of Lakoff, we can see that similar arguments apply. That is, if we tie Murphy & Medin's arguments to this assumed position then we may allow that the different sub-models for *mother* are related by what they call theories. Consider the following example:

> Kim lives with her biological mother, Mary. Mary adopts a child, Jane, to whom Mary behaves in all respects as she does to Kim. Jane calls Mary "mother" and behaves towards her in the same way Kim does. Initially, Kim cannot accept that Mary is Jane's mother but after some time she does.

How is it that Kim's notion of mother appears to change? One possibility is that initially she only has the "biological mother" concept of mother and that she somehow acquires another concept of mother, the "adoptive mother" concept. Even though Kim starts off with the notion of mothers as biological mothers, one can envisage that Kim's theories of the world allow her to construct an explanation of how Jane can call Mary "mother". Indeed, she may come up with a new notion of mother, a mother who is like a biological mother in all respects save the biological relations. That is, Kim, with her notion of a biological mother, equipped only with her theories of the world, can construct the notion of an adoptive mother. The fact that this seems so clearly to be the case undermines the starting point of weak sense selection. Namely, that there may be independent lexical concepts for *mother*. The fact that they may be related by our theories of the world indicates that they are not independent.

We have seen that to posit independent lexical entries is to undermine those very arguments that Murphy & Medin advance in favour of theories. Yet, we take those arguments as very good grounds to accept the theory-based account of concepts. Consequently, we also take it that what is needed to avoid the pitfalls of the sense selection accounts is an extension of this theory-based view. So, although we crucially need a way of describing the difference between lions and stone lions and mothers and adoptive mothers, we need a way of avoiding making an unprincipled distinction at the level of lexical entries.

There are several possible responses to this difficulty. One is to place constraints on the way theories may relate these independent lexical concepts. That is, we may want to place constraints on the possible extensions of any two of the lexical concepts for *mother*, such that one cannot be extended to form the other. However, this flies in the face of the observation that we can do just this. Kim *can* construct a notion of adoptive mother from her notion of biological mothers. So for this to be a viable option, more justification is needed. Another option is simply to reject the arguments of Murphy & Medin that the coherence of senses is determined by theories. The fact that biological and adoptive mothers may form a coherent complex concept has nothing to do with the theories that relate them. As stated earlier we agree with Murphy & Medin's general position. We believe that the only other alternative is to reject the assumptions on which weak sense selection is based. As a consequence, we can see that the interpretation we placed on Lakoff in order to discuss coherence must be rejected. Accordingly, we can only view Lakoff's sub-models as describing something other than lexical concepts.

4 Sense generation

As we have shown, we can characterise lexical entries according to the assumptions of sense selection. We believe that strong sense selection does not provide a viable characterisation. We have illustrated weak sense selection by choosing to interpret Lakoff in this manner. This position is challenged by the arguments in favour of theories as determinants of coherence. In this section we will present an alternative account of lexical representation that is consistent with such arguments.

4.1 Partiality

From our consideration of the arguments of Murphy & Medin concerning coherence, we require independent lexical concepts only in those cases where it is clear that they are unrelated by the kind of theories they discuss. For example, in the case of *bank*, we would require two lexical concepts, one for *river bank* and one for *money bank*. These, after all, seem unrelated unless we construct some ad hoc explanation. In the case of *mother*, though, we would not want to hypothesise independent lexical concepts for "foster mother", "step-mother", etc. Instead we would want to represent *mother* with a single lexical concept. In order for this concept to apply in the cases of all types of mother, that is to capture the phenomenon of semantic flexibility, we would seemingly require it to be general with respect to these cases. As we indicated before, this appears to disallow the possibility of capturing the phenomenon of specificity. However, this conclusion is tied to the assumption that lexical concepts and senses are total objects. Under this assumption, the sense for *mother* cannot be extended or elaborated. That is, we can only have one sense for *mother*. However, that there *are* many different senses is undeniable.

An alternative is to view the various senses as extensions of a single lexical concept, a view that is related to that of Dunbar (1988). That is, senses are related to lexical concepts via a relation of concept [2]. Returning to our example of Rudy at the ballet, the different meanings we may associate with *delightful* can be ordered in the following way. The usher's meaning is only as specific as the lexical concept, and this meaning is compatible with any of the possible extensions of the lexical concept. Ron's meaning, however, is more specific. It doesn't just indicate that there is something that is delightful, it also indicates what it is that is delightful, namely the dancer. Thus, for Ron, the information conveyed by *delightful* would appear to contain the information gleaned by the usher. Similarly, for Reg and Rudy with respect to the usher. Thus, the ordering is one of informational containment. However, if we compare the information conveyed by *delightful* from Ron's perspective with that conveyed from Reg's perspective, we cannot say that either is more informative. Thus the relation of informational containment is undefined for these two senses. It follows that the relation of informational containment is a partial one[3].

Our position then is that semantic flexibility can only be properly captured by considering the lexical concept as a partial object. This can be differently extended according to the interaction of various constraints on each occasion of use. We see such constraints as being provided by the discourse, the situation in which it resides and cognitive models which pertain to both the discourse and the situation. It will emerge that these are the very constraints that underpin the generation of a coherent sense. That being the case, we will see that coherence and flexibility are but two sides of the same coin.

4.2 Coherence in sense generation

We have so far considered the coherence of concepts from the point of view of Murphy & Medin's arguments. On their view coherence is intimately bound to metaphysical issues. As we have seen there are a number of problems associated with this approach. In what follows we will be concerned with the coherence not of concepts but of senses. In this way we avoid being committed to a necessary connection between coherence and metaphysics.

Taking lexical concepts as partial objects seems to provide us with too many degrees of freedom. So, for instance, we seem to require a constraint to prevent certain co-occurrences of features. Consider our lexical concept for *balloon*. We might want to ascribe to it the property of not being able to float in the air. Under the unconstrained account of the extension of partial objects, this is perfectly permissible. However, following on from the arguments of Murphy & Medin, we would want to claim that such an ascription could only be valid if licensed by a theory. In this case, the balloon could be filled with a heavy gas. Hence, in order to constrain what extensions we allow, the process of extending a lexical concept to form a sense must be the very process by which that sense becomes coherent. That is, the process of extension must realise the kinds of theory envisaged by Murphy & Medin.

[2] From here on we use *concept extension* to mean that relation between a lexical concept and the same lexical concept some of whose features have been modified. Modification here include the cases of simple adding of features as well as the denial of features. In no sense do we mean *extension* to refer to the set of objects described by the lexical concept.

[3] Since the relation is partial, it follows that concept extension need not result in a sense that is more specific than the lexical concept.

In order to illustrate the kinds of issues which emerge with respect to coherence in sense generation, we will consider the example of the stone lion, of section 3.2. According to our view, a given sense of a word derives from the corresponding lexical concept and may or may not be an extension of it. Consider describing a stone lion as a lion. We can account for this sense of lion in either of two ways. One possibility is that the lexical concept is general and no extension is required. From our preceding arguments this option would disallow the possibility of simultaneously satisfying the conditions of semantic flexibility and specificity. The only option then is to consider that the lexical concept is specific in the following sense: it contains, or may inherit, among others, feature descriptions such as "animate", "four-legged" and "golden-colour". This will capture the sort of specificity we require for some occasions of use, for example, where we are describing a lion at the zoo. There may, however, be circumstances in which a lexical concept would be extended further, perhaps in order to distinguish between lions and lionesses. More importantly though, as it stands, the lexical concept cannot meet the flexibility requirement. In order to resolve these problems, we will invoke a more general mechanism for concept extension. So, in the case of the stone lion being described as a lion, seemingly the feature "animate" present in the lexical concept is denied in the generated sense. Accordingly, what we require of the constraints which underpin concept extension is that they allow this sort of defeasibility. In section 5 the two views we present will adopt different methods of achieving this.

In principle, we allow that any property ascription provided by a lexical concept can be defeated. As a consequence, we cannot regard the various entities described by a given word as forming a category which has some metaphysical significance. That is, such categories, in contrast to those envisaged by Murphy & Medin, do not "carve the world at its joints". Returning to the example, since a stone lion is inanimate, we require that the extension of "lion" involves at least the defeat of the property ascription concerning animacy. In this way we can meet the requirement of semantic flexibility. Given the lexical concept we have for *lion*, there are circumstances in which it seemingly can't be used. However, via our defeasible mechanism for concept extension, we can override any feature which would prohibit its use. The important point here is that this overriding can only be licensed by theories, *ad hoc* or otherwise.

The mechanism as we have described it so far is a little simplistic. Consider again the case of the stone lion being talked of as a lion. In one sense, we may agree that the statue is a lion but in another we may protest that it is not. That is, we require some choice in the mechanism as spelled out. If we want to affirm that the statue is a lion then we will adopt the defeasible method of extension already outlined. If, however, we want to deny that the statue is a lion we will not extend the lexical concept in this defeasible way. Of course, we may allow the statue to be described as just that, a statue. Again, in such a case we wouldn't require any defeasible extension of the lexical concept for statue.

These different options for concept extension lead to different property ascriptions, that is, different descriptions concerning the same entity. We take these descriptions of an entity to define different perspectives. For a particular cognitive agent, a perspective associates descriptions with given entities. So, if Fred approaches our stone lion adopting the "real lion" perspective, one that is defined by the description given by the lexical concept for lion, he won't categorise it as a lion. However, were he to adopt the "statue of a lion" perspective, one given by a description resulting from defeasible extension of the lexical concept for lion,

he would categorise it as a lion. He could of course adopt a third perspective, the "stone statue" perspective, resulting from a non-defeasible extension of the lexical concept for statue. From this perspective, he would similarly categorise the stone lion as a statue.

We can now see that an adequate characterisation of coherence is not going to be provided solely by the theories which Murphy & Medin discuss. Instead, we will pursue the idea that theories may be local: that is, provided by ad hoc explanations, situational factors and informational requirements. If Fred were to exclaim, "This is not a lion", then we can only ascribe to him the perspective we labelled the "real lion" perspective. It is only this perspective that makes this use of lion coherent. If we attribute the "statue of a lion" perspective to Fred, then this particular use of lion would be rendered incoherent. It follows then that the coherence of a sense can only be defined relative to some perspective. So, clearly, perspectives may involve local or ad hoc theories: we don't carry around our theoretical luggage everywhere we go.

There are many constraints on concept extension that give rise to coherence. Some are to do with the extra-linguistic aspects of word use, such as the agents involved, the situations in which they find themselves and the particular informational requirements they have. Others are to do with the linguistic context, concept combination, for example. In what follows we will focus on the notion of informational requirement and examples of concept combination.

Consider the following puzzle taken from Braisby (1989):

> Fred is sitting on a park bench in London. He knows that at the other end of the park there is a statue of a lion. A schoolgirl approaches him and, explaining that she has been given an assignment to sketch a lion, she asks Fred if he has seen one. Fred replies that he has and points her towards the statue. A little later an exhausted zoo-keeper appears and, explaining to Fred that a lion has escaped from the zoo, he also asks Fred if he has seen one. Fred replies that he hasn't.

Related to our discussion of perspectives, we would like to say that Fred is adopting different perspectives for different occasions. The reasons for choosing the particular perspective he does are to do with the nature of the information required of him. Fred can infer that the schoolgirl requires information germane to lions that one can draw and so adopts the relevant perspective in order to reply in the way that he does. The zoo-keeper, however, places a different informational requirement on Fred. It is this that similarly helps determine the perspective Fred adopts and so allows him to reply as he does to the zoo-keeper.

We would like to distinguish two aspects to informational requirement. One is appropriate to the degree of specificity a given sense may have. So, in the above example, in replying to the zoo-keeper, Fred means "real lion" by *lion*, a meaning given by the lexical concept for lion. In replying to the schoolgirl, however, Fred has a more specific meaning in mind, namely *lion* as meaning "statue of a lion", a meaning that can be derived via defeasible extension from the lexical concept for lion. That is, the informational requirements of the agents involved help determine the degree of specificity of the senses of Fred's use of *lion*. In addition, these requirements lead to a difference in the flexibility of these meanings. That is, the meaning of Fred's use of *lion* in replying to the schoolgirl, is only flexible

enough to describe statues of lions. The meaning of his use of "lion" in replying to the zoo-keeper, however, is different in terms of flexibility: it is only flexible to the degree that it can describe real lions. So the informational requirements of the agents involved can also lead to differences in the flexibility of meaning.

The other constraint on concept extension that we will consider is that provided by concept combination. In the next section we will expound on this within the general framework developed so far. We will do so via the application of two alternative views.

5 Partiality, coherence and concept combination

Our preceding arguments lead to a set of conclusions that we take as requirements on any theory of lexical representation. First and foremost is the commitment to sense generation, in which non-ambiguous words are represented by a single lexical entry, which contributes one lexical concept, and particular senses are generated from this by concept extension. Further, we take the lexical concept to contain a description of the central properties of entities which can be described by the corresponding word. Accordingly, the lexical concept is not general in the Quinian sense; that is, in itself it can only describe a subset of the range of entities which the word itself can describe. Quine's notion of generality would maintain that the lexical concept can describe all individuals the word is used to describe. Entities outside of this subset, but which nonetheless can be described by the word, require a different mechanism. The mechanism we propose is one of concept extension. We see this as being determined by theories, though not ones that carry metaphysical assumptions, *a la* Murphy & Medin. As a consequence, the extensions of concepts are senses, whose coherence is provided by the nature of the extension. The kinds of theories which underpin the generation of senses are those that allow for defeasibility. Similar to the "lion" case we discussed earlier, we will treat such extensions as defining perspectives, and accordingly coherence obtains relative to a perspective. Perspectives themselves are constrained by factors of the situation, local context and the informational requirements of the agents involved. The following two views are based upon these conclusions.

5.1 View 1

This section outlines a possibility for lexical representation developed by Braisby (1989). Braisby refers both to an implementation in C-prolog and also a situation-theoretic interpretation of this implementation. For purposes of brevity, we will not go into these matters here. What we will present is necessarily a simplification. However, the main points will still emerge. In that account a word meaning (or WORM) can be described in a similar way to that offered by the standard classical view. For example, the WORM for *lion* could be described as follows.

$$\begin{bmatrix} \text{animacy:animate} \\ \text{colour:golden} \\ \text{legs:4} \end{bmatrix} \quad (1)$$

When Fred responds to the zoo-keeper, what he is claiming is that he hasn't seen an individual which this description satisfies. But this WORM does not help us in the case of the schoolgirl. For Fred has not seen *any* individual which can be described by this meaning. How then can he reply to the schoolgirl that he has seen a lion? In Braisby's (1989) account,

this case implicates what are called Combinations of sc worms (or COWORMs). COWORMs can also be involved in word meaning. Indeed, a word's meaning may be given by the WORM which corresponds to it or by any one of a number of COWORMs. A COWORM is arrived at by combining a number of WORMs. Taking the "lion" case as an example, there is a combination of the WORMs for *lion* and *statue* as follows.

$$\begin{array}{l} \text{WORM}(\text{LION}, P_1) \\ \text{WORM}(\text{STATUE}, P_1, P_2) \end{array} \tag{2}$$

What the COWORM does is to relate various data structures. It does this by virtue of the fact that "statue" relates two types of data structures. That is, we may view it as an operator. Given a data structure of one particular type, we obtain a data structure of a different type. For example, this particular COWORM, relates the data structure above to one like the following.

$$\begin{bmatrix} \text{animacy:inanimate} \\ \text{colour:\_} \\ \text{legs:4} \end{bmatrix} \tag{3}$$

Here the "statue" WORM is being used to form a different kind of data structure. In particular, it tells us that a statue of a lion is inanimate and does not inherit the colour it has from the thing it represents. The data structure given by this COWORM describes perfectly what Fred *has* seen. Thus, Fred's reply to the schoolgirl is based on this (or a related) COWORM. The crucial aspect of this account is that it is COWORMs which involve a given WORM that may underly the meaning of that WORM's corresponding word. It is given that this is a condition on word use. The particular COWORM that is adopted by a speaker or hearer depends on a large number of factors, ones that have been discussed in the earlier parts of this paper. It is clear in this case though that it is the nature of the information sought from Fred that determines the particular WORMs and COWORMs he employs. This is not unlike the case of Rudy.

Let us suppose that each of our excited ballet lovers has the same WORM for the ballerina as follows.

$$\begin{bmatrix} \text{body:} \begin{bmatrix} \text{shape:fine} \end{bmatrix} \\ \text{hair:} \begin{bmatrix} \text{colour:brown} \end{bmatrix} \\ \text{dance:} \begin{bmatrix} \text{move:pirouette} \end{bmatrix} \\ \text{personality:} \begin{bmatrix} \text{temperament:quiet} \end{bmatrix} \\ \vdots \end{bmatrix} \tag{4}$$

Here a subtlety in the process of combining WORMs is required. When "is delightful" is predicated of the ballerina by Rudy, his friends may do a number of things with the WORM that describes her. One option, that chosen by Reg, is to combine his WORM for the ballerina with that for *delightful* so the feature of delightfulness is added to that aspect of the WORM for ballerina concerning her dance as follows.

$$\left[\begin{array}{l} \text{body:}\left[\ \text{shape:fine}\ \right] \\ \\ \text{hair:}\left[\ \text{colour:brown}\ \right] \\ \\ \text{dance:}\left[\begin{array}{l}\text{move:pirouette}\\ \text{is:delightful}\end{array}\right] \\ \\ \text{personality:}\left[\ \text{temperament:quiet}\ \right] \\ \vdots \end{array}\right] \tag{5}$$

In a similar fashion, Ron combines his WORMs for the ballerina and *delightful*, resulting in a data structure in which "delightful" is a feature of the body attribute of the WORM for the ballerina. Similarly for Rudy. How, though, are we to capture the usher's interpretation?

The usher has a meaning for the ballerina that is the same as our ballet lovers, let us say. However, she has many possibilities open to her. She may combine her WORM of *delightful* with that for the ballerina in any number of ways. The number of options open to her are determined in two ways. Firstly, by the number of complex attributes of the WORM for the ballerina. Secondly, by the number of additional WORMs that she could employ. Similar to the "lion" case where Fred used the additional WORM of "statue" according to the appropriateness of the circumstances, so the usher has similar options too. If, for example, the usher is aware that certain members of the audience play all manner of language-games, then the meaning of *delightful* is simply indeterminate. That is, there are many possible COWORMs that we may take to define these games and the usher has no way of deciding which COWORM to choose. It is this indeterminacy that renders her unable to say what Rudy means.

Our other example involves Kim, Mary and Jane and the problems Kim has when Jane comes to stay. In particular, we want to say that Kim has a notion of what it is to be a biological mother but not an adoptive mother. However, she must come to acquire such a notion. The crucial aspect of the general discussion has been that we do not want to conclude that Kim ends up with more than one lexical entry for *mother*. Again, we will use the mechanism of COWORMs.

We can describe Kim's WORM for *mother* as follows.

$$\left[\begin{array}{l} \text{gender:female} \\ \text{genetics:gen} \\ \text{child:}\left[\ \text{genetics:gen}\ \right] \end{array}\right] \tag{6}$$

How is it that given this WORM for *mother* Kim can later accept Jane's utterances concerning Mary? Again, we will claim that Kim learns of "adopting", which expresses a relation that Kim can utilise in forming the following COWORM.

$$\begin{array}{l} \text{WORM(MOTHER, } P_1\text{)} \\ \text{WORM(ADOPTIVE, } P_1\text{, } P_2\text{)} \end{array} \tag{7}$$

This COWORM will give a new kind of data structure, of a type related by the WORM for *adoptive*. This particular WORM acts to deny the genetic connection between parents and children. Accordingly, we get the following data structure which allows Kim to make sense of Jane's use of *mother*.

$$\left[\begin{array}{l} \text{gender:female} \\ \text{genetics:gen1} \\ \text{child:} \left[\ \text{genetics:gen2}\ \right] \end{array}\right] \tag{8}$$

Within Braisby's framework WORMs and COWORMs can be taken as defining different perspectives. Thus, we might say that Fred has two perspectives: one where he uses the WORM "lion" in isolation, one where he uses it in combination with the WORM for *statue*. Similarly, what Kim has learned is not a new lexical concept or WORM for "mother" but a new way of Combining WORMs (or a new perspective) and similarly with our long forgotten ballet lovers. In a similar vein, we may envisage "concept combinations" being accommodated in this fashion. Let us take as an example, "stone lion". It seems that we would like to say that, at the very least, "stone" implies inanimate while "lion" implies animate. Such an intuitively appealing view precludes the standard set intersective account. Alternatively, the feature structures which describe the component WORMs will fail to unify. Here, we would expect one or other of the component WORMs to be extended so that unification will succeed between the extended senses. Just as before, a Combination is required. Combining the WORMs for "lion" and "statue" renders a sense for *lion* that implies inanimacy. The feature structure described by this COWORM can then unify with that for "stone". However, just as we may view this as taking a particular perspective on what "stone lion" means, we can imagine that this combination allows for a number of different perspectives to be taken. So, although we may take the "statue of a lion" perspective defined by the COWORM of "lion" and "statue", we could also take the "petrified lion" perspective on "stone lion" which is defined by the combined WORMs of 'lion' and 'petrified'. Given suitable intonation and stress, we might also adopt the perspective which gives the meaning "lion that sells stones" which is what we would want to say in the case of "coal merchant" which means "one who sells coal". And so on.

We have seen that by considering the meaning of a word as being described by different data structures on different occasions of use, we open up the possibility of capturing flexibility and specificity. In addition, coherence is captured by the use of COWORMs which we may think of as embodying atomic, local theories. This is consistent with the objectives of the sense generation position.

5.2 View 2

This section presents some of the central aspects of the view of lexical representation and combination sketched in Franks et. al., (1988), and developed in Franks (1989). The outline given here is a simplification of that work in that it will focus primarily upon concept combination. We will consider the way in which one concept's extension constrains that of the other in a head-modifier relationship. In this way, the combined extensions result in a single sense for the noun phrase. The properties descriptions of the lexical concepts correspond to "Central" properties, derive from conventionalised lay theories.

These include Central-essence properties, which relate to the ontological or functional essence of an entity.

Let us again consider our ballet buffs. Suppose that the ballerina is known to them as "Netty". Suppose further that Rudy's first comment had been:

(1) Netty is a delightful ballet-dancer.

Each of the meanings of Rudy's comment, for the ballet-buffs, can be seen to interpret (1) in different ways. On a simple account, we might say that this combination involves the concept for *ballet-dancer* being extended to incorporate the property of delightful. Hence, we might then say that the interpretation of (1) is (2).

(2) Netty dances ballet delightfully.

This is Reg's interpretation. It is based upon an assumption of Explicit Semantic Attachment, in which the major source of the properties of the sense for a NP is the explicitly given head noun, and the properties contributed by the modifier will be evaluated with respect to this noun (Platts, 1979). In (1), the explicitly attached noun is "ballet-dancer". Franks argues that, in order to capture semantic flexibility we must allow for Implicit Semantic Attachment: that is, agents may evaluate the property of being delightful with respect to a different noun concept. How is this effected? The initial combination of the modifier with the head noun unifies their properties and results in a sense which carries the meaning that Netty, who is a ballet-dancer, is delightful in some unspecified way. This is the sense which is generated by the usher. Her sense is compatible with many different extensions, in which the general property of being delightful is determined as a more specific property by evaluating it with respect to an implicitly attached noun concept. For example, Ron's meaning involves a choice of a noun concept such as "people seen in terms of their bodies", which results in an appropriate specification of being delightful. We then have this sense for (1):

(3) Netty is a ballet-dancer who is delightful to look at.

And Rudy's sense for (1) will similarly involve the selection of another implicitly attached N concept. In these cases the different agents are adopting different perspectives on the ballerina. We noted in 4.2. that a perspective defines a description of an entity. Franks argues that perspectives have two components: the categorising sortal noun or nominal phrase (either implicitly or explicitly attached to the construction which describes the entity), and the range of properties which are used to make the categorisation. In this case, we have perspectives whose major contrast is given by their different covering sortals. In contrast, the perspective of the usher has a less specific range of properties, and no particular sortal noun. Consequently, for the usher, the meaning of Rudy's phrase is indeterminate. The flexibility in perspectives will be further evident in the consideration of *stone lion* and similar combinations.

If combining concepts constrain the extension of the head noun, then they may act as defeaters for some of the properties in the noun's lexical concept. This is exactly what happens in what Franks labels Negating Functional Privative combinations, such as *stone lion* and *wooden banana*. These behave in a similar way to to Negating Privatives Proper, which involve adjectives such as *fake* and *false*. The latter have usually been defined as denying such inferences as that from

(4) This is a fake gun.

to

(5) This is a gun.

Franks argues that this conclusion omits the crucially perspectival nature of such categorisations, which we will now examine.

Consider the combination "fake gun". The lexical concept for *gun* might include such Central-essence properties as "internal mechanism for propelling bullets" and "barrell for directing bullets", and other properties such as "trigger mechanism", "barrell mounted over trigger", "colour and weight of metal", and so forth. The lexical concept for *fake* specifies that it will defeat the Central-essence properties of "gun", and allow the ascription of some of the "appearance" properties of "gun" to the sense for *fake gun.* The modifier, then, acts as a Rebutting defeater for the noun's central-essence properties, that is we can ascribe the negation of the properties to the sense. After this combination, where the modifier has priority, we find that the sense for *fake gun* parallels the usher's for *delightful ballet-dancer.* That is, it includes only the general property that the entity has some of the appearance properties of the head noun. But we cannot say in precisely which way it looks like a real gun. It is consistent with many different extensions of this 'appearance' property, through the selection of a particular implicitly attached noun. This is consistent with our intuition that there are many possible ways for a gun to be fake. An entity of the type described by this sense would be indiscernable from a range of entities of the type of possible implicitly attached noun concepts, which might be labelled "toy gun", "replica gun" or "model gun". This means that we might adopt a categorising perspective on such an entity based on the narrow range of "appearance" properties, with respect to any of these nouns, as well as the head noun. That is, we might categorise it thus:

(6) The fake gun is a gun (or toy/replica/model gun) with respect to appearance.

Since this perspective does not incorporate any Central-essence properties, it could not be generalized to a broader range of properties, and hence the categorization is restricted in its applicability. We will refer to such restricted perspectives as Type II Perspectives.

If there is an informational requirement to make the sense more specific, then our agent will access an implicitly attached noun in order to extend the sense. This choice is crucially related to situational factors. If we know that the entity described by *fake gun* has been used to overcome some intruder in a house, we might then select "replica" as the appropriate noun type. This will be used to specify the detail of the "appearance" property, and the sense will also inherit the Central-essence properties of this noun. So the sense for *fake gun* includes properties to the effect that such an entity is a replica of a gun, which "looks like" a real gun in the way that a replica does. Now, we might adopt two perspectives on the entity described by the noun phrase. The first is the Type II Perspective:

(7) The fake gun is a gun with respect to appearance (as a replica).

The second is a Type I Perspective:

(8) The fake gun is a replica with respect to central-essence properties.

A Type I Perspective is one in which the categorised entity has the central-essence properties of the categorising sortal noun.

It seems that exactly the same kind of process occurs for a combination like *stone lion.* Here, the sense can define two different perspectives. Fred's situational context would facilitate a choice of "statue" rather than "rock-formation" or "ornament" as the implicitly attached noun. The Type II Perspective is:

(9) The stone lion is a lion with respect to shape (as a statue).

This is the perspective of the schoolgirl, which Fred adopts in order to respond as he does. The Type I Perspective is:

(10) The stone lion is a statue with respect to central-essence properties.

The perspective of the zoo-keeper would be a Type I: to satisfy his informational requirements, the entity described by *lion* would need to be a lion with respect to central-essence properties. This clearly cannot be supported by the sense generated for *stone lion.* In order for Fred to reply "yes" to any agent asking about the presence of lions, the perspective of the questioner must incorporate a range of categorising properties which subsumes the range of properties in the sense for *stone lion.* The schoolgirl's perspective satisfies this, but the zoo-keeper's does not. Franks argues that this is a requirement on the coherent use of a categorisation statement.

Our third example concerns Kim's quandary. This might be approached through a consideration of another type of privative combination, the Equivocating Privatives. These include combinations such as *blue orange* and *straight banana*, as well as ones involving adjectives like *apparent* and *alleged.* Equivocators are combinations in which we are unable to state whether the entity described by the noun phrase is in fact a member of the head noun or not. We simply have insufficient information. The modifier Undercuts the head N's Central-essence properties. That is, on current information, we are unable to ascribe these properties to the sense: it does not mean, unlike in the case of Negators, that we can ascribe the negation of these properties to the sense. If our information changed, we might reverse the Undercutting.

Consider the combination "apparent friend". The Central-essence properties of the concept for *friend* might include "loyalty", "affection", "absence of self-seeking", whilst the other properties include "solicitous behaviour", "smiles", "various vague kindnesses". And the Modifier's lexical concept specifies the Undercutting of the head noun's Central-essence properties, and an ascription to the noun phrase sense of some of the non-central-essence properties. Having Undercut the Central-essence properties of the head noun, we are again left with a sense which is indeterminate as regards exactly how the "appearance" properties of the head noun are manifest. After accessing an appropriate implicitly attached noun, such as *deceiver*, the general property is made more specific. This sense then defines a Type II Perspective on an entity described by the noun phrase:

(11) This apparent friend is a friend (or deceiver) with respect to appearance.

In order to decide whether this apparent friend is a real friend, we need to extend our information through some process of inquiry. The outcome of this will determine the appropriate combinations of perspectives. In an Affirmative outcome, the apparent friend is a real friend, thus reversing the Undercutting of the Central-essence properties of the head noun, and allowing for a Type I Perspective:

(12) This apparent friend is a friend with respect to central-essence properties.

In a Privative outcome, the sense defines the same combination of perspectives as in the Negating Privatives case.

Let us return to Kim's quandary. This follows a pattern similar to Equivocators. In trying to understand Jane's use of the word *mother*, through trying to adopt Jane's perspective, Kim will Undercut the central-essence properties of her concept of *mother*. Hence she could use this sense to define Jane's perspective, which is of Type II:

(13) Mary is Jane's mother with respect to behaviour.

This allows her to have some understanding of how Jane can call Mary "mother". However, Kim may learn of "adopting", which enables her to determine that, from Jane's perspective, the sense for *mother* involves a Rebuttal of the central-essence properties of the lexical concept for *mother*. But as in the Negators case, this does not prevent Jane from calling Mary "mother": it simply means that she must adopt a particular type of perspective to do so. As in View 1, this implies that Kim is not learning a new lexical concept for *mother*, as in our interpretation of Lakoff, but that she is able to adopt a new perspective - a Type II Perspective - on her mother as Jane's mother.

On this view, then, the flexibility and specificity of a wide range of combinations are addressed by the twin mechanisms of defeasible extension and implicit semantic attachment. Both are viewed within the context of the perspective-relative nature of categorisation. Accordingly, coherence is relative to the two aspects of a perspective, and both must be appropriate for a coherent categorisation.

6 Conclusions

In this paper, we have outlined a general account of lexical representation and presented two approaches within this account and the way they treat the problems of coherence and concept combination. Moreover, we have demonstrated that both approaches promise an adequate treatment of the crucial phenomena of semantic flexibility and specificity. These approaches have been motivated by arguments concerning coherence in sense selection.

Strong sense selection has unfortunate consequences especially regarding the combinatorial explosion of interpretation associated with complex expressions. We take it that this position is not only unfortunate but untenable. Weak sense selection, which is assumed in a number of linguistic approaches, avoids an explosion of this magnitude. However, we have argued that the position is undermined by a consideration of coherence. We concur with the general position of Murphy & Medin that coherence is determined by theories. As a consequence, we expect the relations between senses to be expressed in any theory of lexical representation. This conclusion is in direct opposition to the starting position of

weak sense selection. Our conclusion is that such a starting position is unprincipled. In order to have a principled account of the phenomena we have discussed, it seems a sense generation account is required.

An important implication of the sense generation approach is that lexical concepts should be regarded as partial. The specificity of meaning we need is brought about by the possibilities of concept extension. The mechanism of concept extension, in allowing for the defeat of properties, also provides for the desired flexibility of meaning. Importantly, the constraints on extension are provided by theories which, as we have seen, underpin the coherence of the generated sense. We have strongly argued that theories are local: they may be provided by the immediate situation. In addition, agents can choose to utilise certain theories in preference to others on the basis of their informational requirements. This at once extricates us from Murphy & Medin's connection between metaphysics and theories. It also, crucially, leads us to the conclusion that an agent adopting different theories in his cognitive activities is entertaining different perspectives. Consequently, the examples which we have employed throughout this paper are essentially perspectival in nature as, indeed, is word meaning.

References

Braisby, Nick. 1989. Situating Word Meaning. Paper presented at the Conference on Situation Theory and its Applications, Asilomar, CA.

Clark, Herbert H. 1983. Making Sense of Nonce Sense. In d'Arcais, G. B. Flores. and Jarvella, Ray J. (eds.). *The Process of Language Understanding.* Chichester: John Wiley and Sons. 297-331.

Dunbar, George L. 1988. The Cognitive Lexicon. PhD. Thesis. Centre for Cognitive Science, University of Edinburgh.

Murphy, Gregory. L. and Douglas. L. Medin. 1985. The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Franks, Bradley. 1989. Concept Combination: Towards an Account of Privatives. Edinburgh Research Paper in Cognitive Science.

Franks, Bradley, Terry Myers, and Scott McGlashan. 1988. Defeasibility in Concept Combination. In *Proceedings of the Tenth Annual Conference of the American Cognitive Science Society*, Montreal.

Lakoff, George. 1987. Cognitive Models and prototype theory. In Neisser, U. (ed.). *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization,* 63-100. Cambridge: Cambridge University Press.

Quine, Willard. V. O. 1960. *Word and Object.* Cambridge, Mass.: MIT.

Platts, Michael. 1979. *Ways of Meaning.* London: Routledge and Kegan Paul.